From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Sweden, 2015

STRUCTURAL & FUNCTIONAL STUDIES OF L-PGDS AND SMPDL3A, ENZYMES IN LIPID SIGNALING FAMILY

Sing Mei LIM

Karolinska Institutet

Nanyang Technological University-Karolinska Institutet Joint Ph.D

Stockholm 2015

Published by Karolinska Institutet.
Printed by E-print AB

ISBN 978-91-7676-170-0

Structural and functional studies of L-PGDS & SMPDL3a, enzymes in lipid signaling family

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Sing Mei LIM

Principal Supervisor:
Pär Nordlund
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Biophysics

Co-supervisor:
Konstantin Pervushin
Nanyang Technological University
School of Biological Sciences
Division Structural Biology and Biochemistry

Opponent:
Andrea Mattevi
University of Pavia
Department of Biology and Biotechnology

Examination Board:
Sven Pettersson
Nanyang Technological University
Lee Kong Cian School of Medicine
&
Karolinska Institutet
Department of Microbiology, Tumor and Cell biology

Günter Schneider
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Structural Biology

Mats Sandgren
Department of Chemistry and Biotechnology
Swedish University of Agricultural Sciences

To my beloved

Would it not be strange if a universe without purpose accidentally

created humans who are so obsessed with purpose?

Sir John Templeton

ABSTRACT

Enzymes are indispensable in maintaining the biological system. They metabolize complex molecules to supply nutrients, to produce energy, to regulate transcription of gene expression, and to control the concentration of effective signaling molecules in a cell, thus maintaining the homeostasis of biological system. This thesis summarizes the study of the structure and function of two enzymes in lipid signaling family using integrative application of X-ray crystallography, solution NMR spectroscopy, light scattering, ITC and thermal shift assay.

Lipocalin prostaglandin D synthase (L-PGDS) is a tissue specific prostaglandin D_2 producing enzyme with a lipocalin fold. Apart from its enzymatic role, it is known to act as a lipophilic ligand carrier. Crystal structure of human L-PGDS and substrate analog altogether with NMR spectroscopy experiments revealed binding sites for substrate catalysis and entry. NMR titration experiments with membrane mimetic showed that L-PGDS has intrinsic membrane binding affinity depending on the ligand bound. These results allowed a model of substrate catalysis and product egression to be proposed, hence, converging the enzymatic and transporter role that has been reported in literature previously. Since prostaglandin D_2 is a pivotal inflammatory signaling molecule, molecular understanding of L-PGDS is important to facilitate future regulation of the prostaglandin isomerase. The dynamics of substrate-product exchange may guide future design of this lipophilic carrier as vehicle for drug delivery.

The second enzyme, human acid sphingomyelinase like 3a (SMPDL3a), belongs to a metallophosphodiesterase family and shares close sequence identity with human acid sphingomyelinase (aSMase). SMPDL3a's structure is reported for the first time revealing its binuclear catalytic core site bound with Zn metal. Even though it was presumed to be part of the lipid hydrolase family, enzymatic assays showed that it metabolizes nucleotides and modified nucleotides like CDP-choline, CDP-ethanolamine and ADP-ribose. Subsequently, CDP-choline soaked crystal revealed 5' cytidine monophosphate (CMP) ligand bound in the catalytic site due to spontaneous catalysis. Its α-phosphate forms key interactions with histidine residues in the binuclear center. Based on this CMP-enzyme structure, general catalytic mechanism of aSMase family can be proposed. Besides, SMPDL3a also serves as a template for aSMase catalytic domain homology modeling. Further study on enzymes in the acid sphingomyelinase family can now be guided by the newly available structural information.

LIST OF SCIENTIFIC PAPERS

I. **Sing Mei Lim**, Dan Chen, Hsiangling Teo, Annette Roos, Anna Elisabet Jansson, Tomas Nyman, Lionel Trésaugues, Konstantin Pervushin, and Pär Nordlund. Structural and dynamic insights into substrate binding and catalysis of human lipocalin prostaglandin D synthase. *The Journal of Lipid Research*. Jun 2013; 54(6):1630-1643.

II. **Sing Mei Lim**, Kit Yeung, Lionel Trésaugues, Hsiangling Teo, and Pär Nordlund. The structure and mechanism of human sphingomyelin phosphodiesterase like 3A, an acid sphingomyelinase homolog - with novel nucleotide hydrolase activity. *Accepted for publication*.

Paper not included in this thesis:

Sing Mei Lim, Kit Yeung, Hsiangling Teo and Pär Nordlund. Structure and mutational analysis of human 15LOX-2 apozenyme provides insights into substrate binding. *Manuscript*.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

L-PGDS	Lipocalin Prostaglandin D synthase
ITC	Isothermal titration calorimetry
MALS	Multiangle Light scattering
HSQC	Heteronuclear Single Quantum coherence
NOESY	Nuclear Overhauser Effect spectroscopy
Sf9	Spodoptera Frugiperda
SMPDL3a	Sphingomyelin phosphodiesterase like 3a
ASMase	Acid sphingomyelinase
IMAC	Immobilization affinity chromatography
SEC	Size exclusion column
TEV	Tobacco Etch Virus
PGH_2	Prostaglandin H_2
MME	Monomethyl ether
SAD	Single-wavelength Anomalous Dispersion
MAD	Multiple-wavelength Anomalous Dispersion
SIR	Single isomorphous replacement
MIR	Multiple isomorphous replacement
RF	Radio frequency
FID	Free induction decay
CSP	Chemical shifts perturbation
DPC	Dodecylphosphocholine
c.m.c	Critical micelle concentration
AA	Arachidonic acid
COX	Cyclooxygenase
LOX	Lipoxygenase
H-PGDS	Hematopoietic prostaglandin D_2 synthase
SA	Substrate analog
PA	Product analog
NPD	Niemann Pick Disease
PAP	Purple acid phosphatase
ATP	Adenosine triphosphate
CDP	Cytidine diphosphate
CMP	Cytidine monophosphate
cAMP	Cyclic adenosine monophosphate
CLP	Calcineurin like phosphodiesterase

1 ENZYMES – CATALYSTS OF THE CHEMISTRY OF LIFE

In the mid-18th century, scientists began to observe that saliva is able to breakdown starch [1], and almost a few decades later, Payen and Persoz discovered and isolated the starch-decomposing compound which they called "diatase", now commonly known as amylase [2]. Twenty years after the release of Payen's discovery, Louis Pasteur proposed that a "vital force" in yeast promotes fermentation of sugar to alcohol in 1858 [3] and Wilhem Kühne used the word "enzyme" to describe the process of fermentation. It was subsequently used to refer to the protein that catalyzes the biochemistry reaction. In fact, the term "catalyst" was coined by Swedish chemist, Jöns Jakob Berzelius in 1835 based on work from other scientists whom he collected and observed. An excerpt of his published statement:

"*It is then shown that several simple and compound bodies, soluble and insoluble, have the property of exercising on other bodies and action very different from chemical affinity. The body effecting the changes does not take part in the reaction and remains unaltered through the reaction." – [4]*

This sets the fundamental foundation for the concept of catalysis. Now, we understand that a catalyst acts to speed up a chemical reaction by lowering its activation energy while remaining unchanged at the end of the reaction (Figure 1). This energy barrier is important to prevent chemical reactions from occurring spontaneously. However, if unaided, most chemical reactions would occur very slowly due to the high activation energy required. Therefore the presence of a catalyst is vital to increase the speed of reactions and allows only certain reaction to take place at a time. These controlled-chemical processes in a biological system are normally governed by enzymes.

At the end of 19th century, many scientists began to draw great interest in the study of enzymes and this have contributed significantly in shaping the current understanding of enzymology and cellular processes where they are involved. Buchner was the first to describe enzymatic reaction outside of living cells and he was awarded the Nobel prize in Chemistry for his discovery of cell-free fermentation [5]. This enabled the application of cell free enzyme systems in food and wine processing during the 20th century.

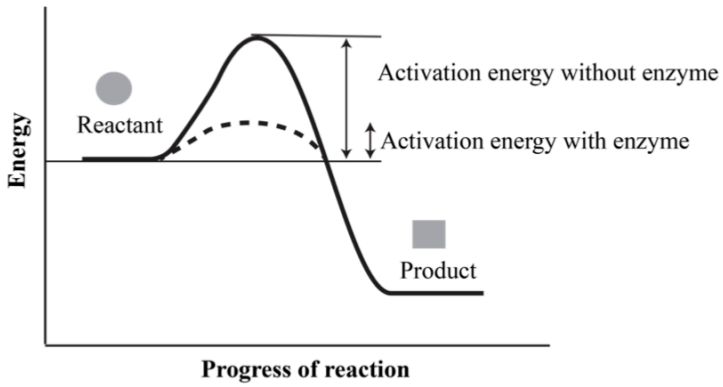


Figure 1: Energy barrier of chemical reactions (also termed activation energy) is lowered in the presence of enzyme.

Around 1890s, scientists began to study the reaction rate of enzymatic induced biochemical reaction when they realize that enzymatic reaction can be observed outside of the living cells. Many deduced that enzyme performs its reaction by forming complex with its substrate and a chemist, Emil Fischer, proposed a lock and key model to explain the enzyme-substrate interaction [6]. Subsequently L. Michaelis and M. Menten proposed the basic idea of enzymatic reaction in the following reaction scheme.

$$E + S \underset{k\text{-}1}{\overset{k1}{\leftrightarrow}} ES \overset{k2}{\rightarrow} E + P$$

For a single transition state, the reaction would result in forming the first intermediate (ES) before proceeding to yield the product (P). The rate of constant k_1 described the binding of substrate to enzyme whereas the k_2 described the catalytic rate of reaction turning from substrate to product. The process can also be reversed with rate constant $k\text{-}_1$. When a reaction reaches steady state, the ES intermediate concentration is assumed to be relatively constant where the rate of ES formation is close to or same as the rate of ES dissociation. This steady state approximation allows derivation of the Michaelis-Menten equation [7].

$$V_0 = \frac{\nu_{max}[S]}{K_M + [S]}$$

The Michaelis-Menten equation explains kinetics of enzymatic reaction. From the equation, the reaction rate dependence on substrate and enzyme concentration can be explained. For example, a low K_M value indicates a strong binding between substrate and enzyme in which small amount of substrate concentration result in 50% of maximum reaction velocity. Enzyme kinetics enables scientists to make important observations such as measuring the change of reaction rate with respect to change of pH and temperature. This pH and temperature dependence is attributed to the protein nature of enzyme but the identity of

enzyme was still unknown at that time. It was not until 1926 when J. B. Sumner showed that the enzyme urease is a protein and it can be isolated and crystallized. Despite confronted with much skepticism, his findings were soon confirmed by J. H. Northrop and W.M. Stanley [8-11].

It is now well established that enzymes are versatile catalysts that can have exquisite selectivity. They facilitate most chemical reactions in biological system including processes controlling gene expression, signal transduction, energy production, molecular motors and cellular homeostasis. The enzyme's function can be assisted by catalytic cofactors such as prosthetic group, metal ions and co-enzyme, or further regulated by posttranslational modifications, interacting proteins or allosteric modulation. Most enzymes are proteins but many proteins are not enzymes. Proteins have three-dimensional structures comprising a chain of amino acids connected by peptide bonds. Apart from proteins, RNA has also being identified with enzymatic ability [12]. Nonetheless, this thesis will focus only on protein enzymes.

1.1 ENZYME STRUCTURE AND FUNCTION

Sumner's urease extraction and crystallization work did not just provided a breakthrough in the understanding of enzymes but it also set the stage for much later work in the field of structural biology. From 1912 to 1930s, there were many small chemical structures being determined using X-ray crystallography. The physical method was introduced into the biological world when Crowfoot, D. first demonstrated that protein crystals showed distinct X-ray diffraction patterns like chemical crystals [13]. Based on fiber diffraction method, J. Watson, F. Crick and R. Franklin first uncovered the structure of DNA nucleic acid [14]. It was an extraordinary breakthrough and an important step towards the formulation of the central dogma of molecular biology that explains the flow of information from DNA to mRNA and subsequently translated to proteins. The excitement of viewing DNA structure spurred scientist like M. Perutz and J.C. Kendrew to work towards the first protein structure – myoglobin in just five years later [15]. It was soon followed by the first enzyme structure, the hen egg white lysozyme, in 1965 [16, 17].

Structural elucidations of enzymes have contributed significantly in establishing the principles of enzyme function. The emergence of serine proteases structures including trypsin, pepsin, chymotrypsin and papain gave early insights into the structural basis for catalysis and substrate specificity of these enzymes [18, 19]. Following the structural information explosion, especially in the last 20 years, the field now provides a comprehensive

view of many important biological systems on the structural level, including enzyme functionalities. Since many drug targets are enzymes, these structural information, and accompanying strategies to rapidly determine protein structure, now serve as a valuable source to guide rational drug design.

1.2 THE LINE BETWEEN SPECIFICITY AND PROMISCUITY

Enzyme has an active site where the substrate is bound and chemical reaction takes place. Fischer's lock and key model in 1894 proposed that the substrate has to be perfectly complimentary to the active site for productive interactions [6]. In this model, the specificity of enzyme is governed by the fact that geometrically incompatible substrate would not be the right 'key' to activate the 'lock' (enzyme) (Figure 2A). In 1958 D.E. Koshland suggested a modification of the prevailing model to the 'induced-fit model' based on his study on hexokinase [20]. In the 'induced-fit model', both the enzyme and substrate were deemed to be moderately flexible and the encounter of enzyme-substrate in the correct orientation would induce conformational change in the active site to promote catalysis (Figure 2B). Scientists gradually disfavor the 'lock-and-key' model when more and more enzyme structures with and without their substrate(s) have been unveiled. These atomic resolution structures showed that enzymes in their substrate-bound and unbound states could adopt different conformations.

The enzyme substrate interactions are often quite specific in order to tightly control the *in vivo* chemical transformations. As a result, enzymes are generally classified by the types of chemical transformation they catalyze which usually include a defined substrate. However, there are also many enzymes which have a relatively broad substrate specificity such as pyruvate decarboxylase [21], pepsin [22] and chymotrypsin [23].The phenomenon where enzyme catalyzes substrate other than the substrate it physiologically specialized is termed substrate promiscuity [24, 25]. Substrate promiscuity can be achieved by for example structural plasticity at the active site where it could adopt alternate conformations [26, 27].

The molecular level of specificity can also be regulated by gene expression, compartmentalization and allosteric modification to impart specificity of action at the system level. For example, many enzymes are not constitutively expressed or always in the active form, instead of protein level, localization and activity of enzyme can be controlled by cellular or external stimulation. Even if a promiscuous enzyme was expressed constitutively, its cellular localization, activation by effectors and availability of substrates set boundaries for

its enzymatic action. Nonetheless, most enzymes are still highly specific and efficient in their role.

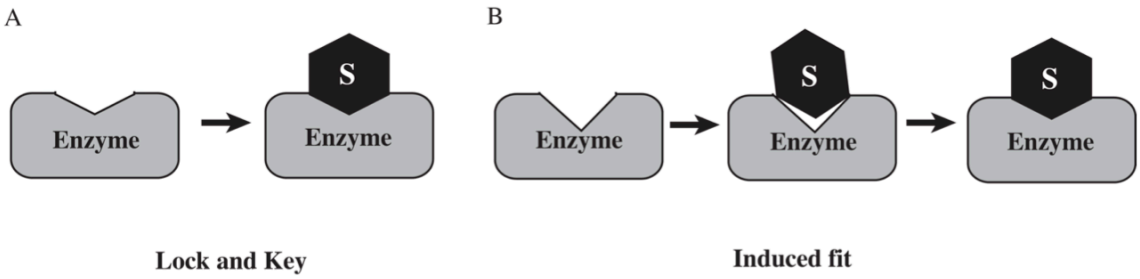


Figure 2. Cartoon depiction of Fischer's lock and key model in (A) versus Koshland's Induced-fit model in (B)

1.3 ENZYMES IN METABOLIC PATHWAYS

Enzymes, given the above-discussed nature and modus operandi, are indispensible and paramount in all types pathways and processes in biological systems. A series of enzymes can be localized to catalyze stepwise biochemical reactions that are more or less interdependent of one another. The arachidonic acid metabolism pathway is one such example. The initial substrate phospholipid is converted to an arachidonic acid (AA) as the first committed step in the pathway. The long hydrocarbon chain of AA can be further oxygenated via three specific enzymes into three distinct signaling pathways. Therefore metabolism of arachidonic acids generates a class of lipid signaling molecules known as eicosanoids. Eicosanoids signaling is further elaborated in Chapter 4 in the study of enzyme lipocalin prostaglandin D synthase (L-PGDS).

The second enzyme in our study is a homolog of the lipid modifying enzyme, acid sphingomyelinase (aSMase), that is involved in ceramide metabolism. Ceramide belongs to a class of lipids known as sphingolipids, which carries a sphingoid backbone and it can be produced via three pathways; de novo synthesis from palmitate [28] and serine, salvage pathway from recycled sphingosines [29] or breakdown of sphingomyelin [30]. Enzyme ASMase produces ceramide by breaking the phosphodiester bond of sphingomyelin. However, our studies and other literature reports found that the conserved catalytic domain of the acid sphingomyelinase like 3a (SMPDL3a) enzyme selects nucleotide substrate instead of sphingomyelin [31]. This is a classical example of enzymes with conserved structural domains but with modified substrate pocket for substrate specificity. From an evolutionary point of view, this is efficient way to develop specialization. Nonetheless, the exact

physiological substrate and biological function of SMPDL3a is unfortunately still unclear. The work in Paper II helps to establish a structural and biochemical framework for future study of this protein and the definition of it´s biological role.

2 AIM OF THESIS

A total of 110,789 protein, DNA and RNA structures were deposited in the Protein Data Bank (www.rcsb.org) as of August 2015. Among those, 98720 were solved with X-ray crystallography, 11077 by nuclear magnetic resonance (NMR) spectroscopy, and 809 by cryo-electron microscopy (cryo-EM) [32]. This exponential growth of structures is fueled by innovative techniques at different stages of the structure elucidation process and high-throughput automation. X-ray crystallography celebrated its 100 years anniversary in 2014 and is still the dominating method in structure biology. Most importantly, together with X-ray crystallography, other structural methods such as NMR, high-resolution cryo-EM and small-angle X-ray scattering (SAXS) can be used in parallel to address a specific structural problem. This idea of integrative structural biology is slowly gaining ground as well as further integrating biophysical, biochemical analysis, cellular studies, bioinformatics and computations biology to give a more holistic view of protein's function and mechanism.

The aim of this thesis is to gain structural and functional insights into two important lipid enzyme families in human cells by integrating several techniques including X-ray crystallography and solution NMR spectroscopy as well as biophysical binding studies and biochemical activity assays.

Paper I

Human lipocalin Prostaglandin D synthase (L-PGDS) has previously been in the spotlight for structural study using X-ray crystallography, NMR spectroscopy and SAXS individually. Despite several mouse L-PGDS structures being reported, these structures usually harbor a catalytic residue mutation with an empty substrate binding pocket and unclear the structural basis of the enzyme and transporter activities. Therefore the objective of this project was to uncover the molecular mechanism of substrate binding, catalysis and product release of wild type L-PGDS utilizing a combination of X-ray crystallography and NMR.

Paper II

Sphingomyelin phosphodiesterase like 3a (SMPDL3a) has been proposed to share similar function with its close homolog acid sphingomyelinase (aSMase), based on the conservation of its metallophosphodiesterase domain. Apart from that, little is known about this protein except its regulation by liver X receptor, cAMP activation and cholesterol loading on macrophages as well as absence of sphingomyelinase activity. This project aims to shed light

on the structure, function and mechanism of SMPDL3a, with possible implications for an understanding of the aSMase family.

3 FROM PROTEIN TO STRUCTURE AND FUNCTION

3.1 RECOMBINANT PROTEIN EXPRESSION

In order to obtain large amount of highly pure and homogenous proteins for structural studies, recombinant protein over-expression is most often an essential requirement. Proteins can be over-expressed in either a cell-based system or a cell free system but currently cell-based expression is still most commonly used. Two types of cell-based systems were used in this work and they will be discussed further in the following sections.

3.1.1 Bacterial expression systems

The bacterial system is the oldest and most widely used cell-based system for recombinant protein expression due to its efficiency, simplicity and economical attributes. Prokaryotic cells have simpler transcription-translation machinery and a fast growth rate . It is genetically easy to manipulate and quick to obtain large cell mass containing proteins of interest. The target gene is cloned into bacteria expression vectors and subsequently the plasmid is transformed into bacteria cells for amplification. Under antibiotic selection pressure specific to the recombinant plasmid, only bacteria cells with the target gene is multiplied and subsequently induced to express the heterologous gene. The bacteria system can be used to also express various types of isotope labeled proteins for NMR experiments.

3.1.1.1 L-PGDS recombinant protein expression

In Paper I, the human lipocalin prostaglandin D synthase cDNA was cloned into pNIC-CH3 vector containing a C-terminal non-cleavable hexa-His tag [33]. *Escherichia coli* (E.coli) of BL21 (DE3) Rosetta T1R strain was transformed with the vector containing the gene of interest, then cultured in Terrific-broth (TB) media. The production of L-PGDS was induced using isopropyl-1-thio-β-D-galactopyranoside (IPTG) and harvested after 16-hours growth at 37°C.

Isotope-labeled proteins for NMR studies were expressed similarly in M9 minimal media supplemented with ^{15}N-labelled ammonium chloride (HN_4Cl) for single labeled protein. Doubled labeled protein was expressed with both ^{15}N-HN_4Cl and ^{13}C-labeled glucose added. The M9 media was also supplemented with ^{15}N-HN_4Cl and ^{13}C-labelled leucine and alanine to label the respective residue type in a non-deuterated environment.

3.1.2 Insect cells system

Another increasingly popular cell-based system is the binary baculovirus-insect cells system. Baculovirus is referred to a family of virus with large, circular double stranded DNA that has a rather specific host range. It is known to infect primarily anthropods like insects [34]. Years of development have led to a more optimized Baculovirus genome for recombinant expression of foreign protein in insect cells. A commercially available Bac-to-Bac method based on Luckow et al's paper is adopted in our production of human acid sphingomyelinase like protein (SMPDL3a) in Paper II [35].

This system employs a baculovirus shuttle vector known as bacmid. The gene of interest is first cloned into a vector and transformed into *DH10Bac* competent cells that carry a bacmid with transposition insertion site (attTn7) and a transposition helper plasmid. This allows the transposition of gene of interest from the vector to the bacmid which disrupts the lacZα reading frame, thus making the cells unable to produce β-galactosidase. These cells form only white colonies on blue-gal containing agar (Figure 3A) [35]. The recombinant bacmids are then isolated and validated with PCR (Figure 3B) before transfection.

Common insect cell strains selected for Baculovirus propagation are Sf9 or Sf21 cells derived from the pupal ovarian tissue of *Spodoptera frugiperda* worm and High-Five™ cells from cabbage looper [34]. They can be grown either as suspension cells or adherent cells at 27°C. Transfection of recombinant bacmid by transfecting agents like cellfectin induced the insect cells to produce baculovirus that express the heterologous gene. After 72 hours of infection, the insect cells will lysed, releasing viruses into the media, which is known as P0 virus stock. The P0 virus stock is used for subsequent infection and propagation to P1 and P2 viruses for large-scale (volume) protein production.

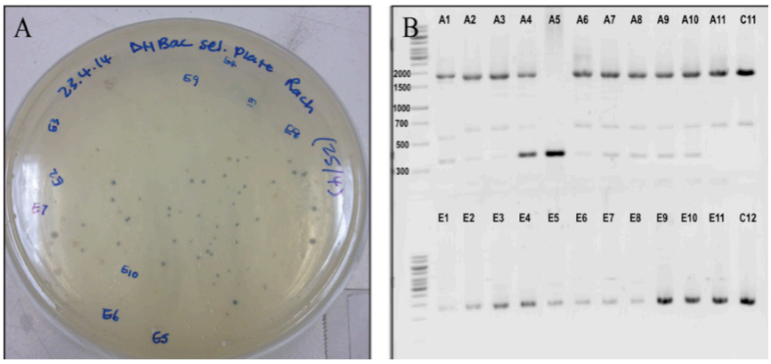


Figure 3: (A) Blue white colony screening after DH10Bac transposition. (B) PCR validation of white colonies extracted plasmids to verify the size of transposed gene of interest.

3.1.2.1 SMPDL3a protein expression

The gene that encodes SMPDL3a was cloned into pFastBac-Sec vector. This vector includes a secretory signal peptide in order to express target gene as a secreted protein with an N-terminal cleavable His tag for affinity purification. Bacmid carrying SMPDL3a gene is transfected in Sf9 cells for protein expression. Since SMPDL3a is produced as a secreted protein, the supernatant was collected for subsequent protein purification.

The baculovirus insect cell system is especially beneficial for SMPDL3a protein as it requires post-translational modification like N-linked glycosylation. However insect cells are incapable of complex glycosylation as they are able to form only simple oligo-mannose sugar chain [36]. This may affect the biological function of the enzyme if complex glycans are essential for its *in vivo* activity.

3.1.3 Protein purification

After protein over-expression, proper purification is necessary and critical for structural studies to ensure that the protein is soluble (except for solid-state NMR), pure, homogenous, and in good amount for structural characterization. This is usually achieved by chromatography where the protein is separated based on selective binding, isoelectric point, hydrophobicity or size.

3.1.3.1 Affinity tag purification

Affinity chromatography is a common and effective method to separate target protein from protein mixture based on selective interaction with a ligand. The ligand that is specific towards target protein is coated on a stationery phase and will bind primarily to the target protein when a protein mixture (mobile phase) is passed through. Proteins expressed with the hexa-histidine fusion tag have high affinity towards nickel beads and this strategy is often used for initial purification of recombinant proteins. A nickel-NTA beads slurry packed in a column can act as the stationery phase to selectively bind the His-tagged protein. Proteins that do not bind specifically to the Ni-NTA will flow through the column upon washing. His-tagged protein can be recovered by increasing the imidazole concentration or by altering the pH of the washing buffer (mobile phase). Imidazole can compete for interactions with nickel NTA while changing the pH alters the protonation state of imidazole nitrogen on histidine ($pK_a = 6.0$) and disrupts the histidine-nickel interaction.

Both L-PGDS and SMPDL3a were expressed with His-tag fusions but purified in a slightly different manner. As L-PGDS protein is expressed in bacteria cells, it is necessary to lyse the

bacteria cell wall by sonication and/or other disruption methods to extract the protein before chromatography. Meanwhile SMPDL3a was secreted into the media, hence no cell lysis was required. The media containing target protein was collected for purification. However, since the histidine-nickel interaction is sensitive to pH, media was first titrated with Tris pH 8 buffer to adjust the slightly acidic pH to a neutral pH. After affinity purification, SMPDL3a was proteolysed by Tobacco Etch Virus (TEV) to remove the His-tag. In some cases, host proteins that are rich in histidines can be isolated together with the target protein. Hence, a second step of chromatography is required to ensure the purity.

3.1.3.2 Size exclusion chromatography (SEC)

This method separates molecules in solution by size (molecular weight). In this case the stationary phase consists of inert porous matrix that retards smaller proteins while allowing larger proteins to flow through, separating the mixture in an isocratic manner. Both L-PGDS and SMPDL3a were subjected to size exclusion chromatography. An interesting observation was that L-PGDS protein elution fractions appearred yellow but the color reduced with increasing elution time. The yellow fraction and colorless fractions were separated for crystallization and NMR experiments initially. No crystals were formed from the yellow fraction and it gave poor NMR spectra (detailed discussion in *Section 3.6.2*) hence only the colorless fractions of L-PGDS were pooled and used for all subsequent crystallization and NMR measurements.

3.2 THERMAL STABILITY ASSAY

3.2.1 Thermal aggregation method

All proteins are stable within a specific pH and temperature range where its structural integrity is preserved to allow it to perform its function. Thermal induced denaturation of proteins has been a vital tool to access protein conformation stability. Binding of ligands such as co-factors, effectors, products or inhibitors often increase protein stability, but sometimes decrease stability (Figure 4). Ligands that induced significant stabilization have been shown to increase the probability of forming protein-

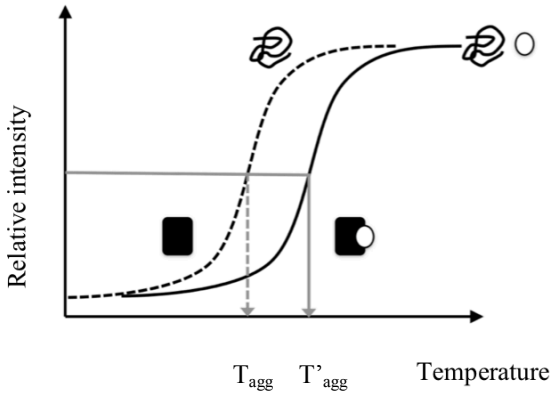


Figure 4: Protein stabilization curve in the presence and absence of ligands.

ligand complexes in crystal structures [37, 38]. Measurement of protein thermal stability can be done in a various ways [37, 39]. Differential scanning fluorometry that adopts SYPRO orange dye to bind to hydrophobic core of protein upon unfolding is not suitable for proteins with hydrophobic patches like L-PGDS. Therefore, in Paper I, a Harbinger Technology, Stargazer-384™ instrument is employed, it uses the concept that when heat denatured proteins are unfold they often aggregate and these aggregates can be detected by light scattering. This approach constitutes a convenient and high-throughput method to screen for ligands and compounds suitable for protein stabilization and ligand binding. The stargazer method is used to investigate whether different compounds were ligands for purified L-PGDS and as a guide for strategizing complexes generation for crystallographic studies. Substrate analog (SA U44069) was shown to give strongest stabilization and thus was used for the co-crystallization experiments that eventually gave the enzyme-SA complex structure.

3.2.2 Dot-blot method

The stargazer method requires relatively high concentrations of proteins and it has a maximum temperature detection of about 80°C. Modified detection method based on similar principle that has been discussed in *Section 3.2.1* can be used to determine thermal stability of protein that are less soluble or with aggregation temperature above 80°C. Instead of monitoring the aggregated proteins, the remaining soluble proteins are detected in a dot-blot after a centrifugation or filtration step. In practice, the protein sample is heated in a PCR iCycler followed by high-speed centrifugation to pellet aggregated protein. Aliquots of supernatant from each temperature point can be taken for SDS PAGE analysis (Figure 5). In order to facilitate large sample size analysis, aliquots are dotted on nitrocellulous membrane instead and subsequently probed with a protein-specific antibody, or in the case of His-tagged SMPDL3a, a His-probe antibody (Figure 5). Relative intensity of coomassie stained band or Western blot signal is measured by ImageJ/ImageLab software. Normalized relative intensity is plotted against each temperature point and fitted with non-linear regression in Graphpad. This was useful for the analysis of SMPDL3a where its native aggregation temperature is approximately 80°C and further stabilization by ligands and metals was not possible to be measured by Stargazer instrument at these high temperatures.

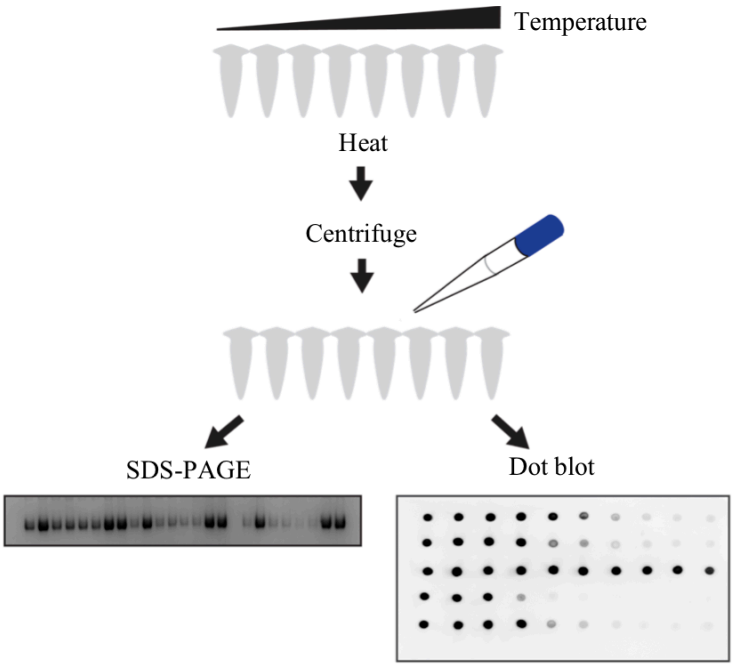


Figure 5: Cartoon description of thermal stability assay for protein SMPDL3a.

3.3 ISOTHERMAL TITRATION CALORIMETRY (ITC)

Isothermal titration calorimetry (ITC) can be utilized to characterize the thermodynamics of enzyme-ligand binding interactions. Since chemical reactions and binding is accompanied by change in enthalpy, measuring the heat released (exothermic) or gained (endothermic) during enzymatic reaction can be correlated to the rate of the reaction [40]. In the case where non-substrate ligand is supplied to the system, ITC can measure the affinity and thermodynamic parameters for binding instead. The experiment setup of MicroCal iTC200 consists of a reference cell for water, a sample cell for protein sample and a syringe for titrant. A series of incremental injections of the titrant are made within specific time intervals throughout the course of the titration experiment. For each injection, the change of enthalpy is measured and plotted against the concentration of ligands introduced into the protein system. Binding affinity and occupancy of ligand(s) can subsequently be determined. Due to the high reactivity of L-PGDS substrate, PGH_2, it is challenging to measure the binding affinity of enzyme-substrate. Therefore the more stable product, PGD_2 was used as a titrant in the study.

3.4 MULTI-ANGLE LIGHT SCATTERING (MALS)

Multi angle light scattering uses the principle of Rayleigh scattering to assess protein homogeneity, molar mass and oligomerization state. Polarizable macromolecules in solution are able to scatter light. The intensity of scattered light is proportional to the concentration of macromolecules and it is enhanced when oligomer is present because oligomer contributes to scattering in coherent phase. Wyatt Technology MALS system is coupled with a size exclusion column (Superdex 200 5/150 GL) and consists of three detectors: absorbance detector (UV), a static light scattering detector and a refractive index detector. MALS was used to determine to oligomeric state of SMPDL3a and the experiment revealed a monodisperse peak revealing that SMPDL3a exists as a monomer in solution (Figure 6).

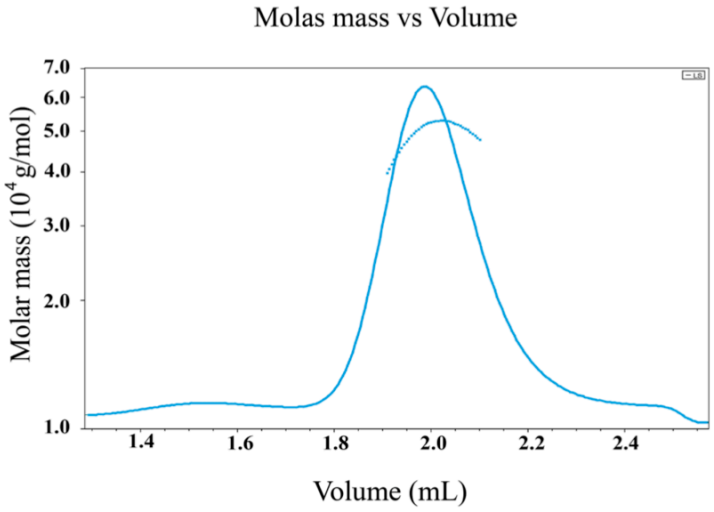


Figure 6: MALS profile for SMPDL3a protein in 20 mM HEPES pH7.5, 300 mM NaCl and 10% glycerol. Sample is monomeric and homogeneous eluting at mass 50.6 g/mol.

3.5 MACROMOLECULAR X-RAY CRYSTALLOGRAPHY

As discussed above, X-ray crystallography is the most commonly used method for elucidating molecular structure of nucleic acid and proteins. The foundation of the method is based on the scattering property of light whereby diffraction of radiation with wavelength in atomic range can resolve atomic structures of macromolecules, if X-ray radiation of sufficiently low wavelength is used, typically in the order of 0.1 nm. In 1913, W.L. Bragg and W.H. Bragg proposed Braggs's Law to describe the scattering properties of X-rays when they encounter atoms organized in an ordered lattice (crystal). This phenomenon termed X-ray diffraction can be used to determine the atomic structure of the molecule(s) in the ordered crystal.

3.5.1 Principles of X-ray diffraction

Bragg's Law states that when two in-phase X-rays incident beams are scattered by atoms that are separated by constant distance d in a crystal lattice, the path differences between the two waves is $2dsin\theta$ [41] (Figure 7). This constructive interference of two scattering waves reflected by the consecutive lattice plane will produce an intense peak when measured with an X-ray detector.

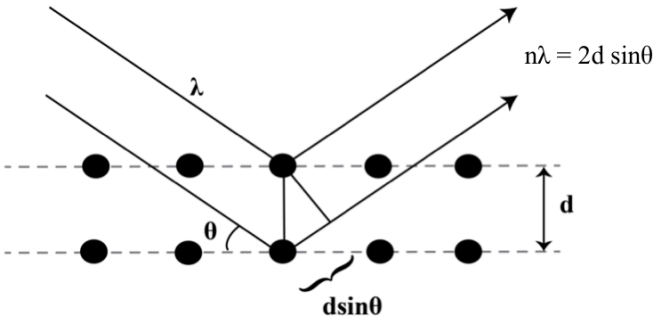


Figure 7: Depiction of Braggs Law.

Bragg's Law provides the location of diffraction peaks corresponding to the atom arrangement but measurements of the diffraction peaks' intensities are required to calculate the electron density distribution in order to construct the protein structure. The following Fourier transform describes how this electron density distribution relates to the strength of the diffraction peak.

$$F(hkl) = \iiint_{x,y,z=0}^{1} \rho\,(x,y,z)\, e^{2\pi i\,(hx+ky+lz)}\, dxdydz$$

F (hkl) is referred as structure factor and the reverse Fourier transform of the above equation allows for the calculation of the electron density distribution in a unit cell,

$$\rho(x,y,z) = \frac{1}{V}\sum_{h,k,l} F(hkl)e^{2\pi i\,(hx+ky+lz)}$$

and,

$$F\,(hkl) = |F(hkl)|e^{i\alpha}$$

The structure factor amplitude |F(*hkl*)| can be measured directly from the intensities of diffraction peaks but the phase α has to be obtained indirectly. The later is done either by additional experiments where site specific distortions are generated in the structure leading to small changes in diffraction intensities, or by using a homologues structure to generate initial phase information. When the electron density map is of sufficient quality, the atomic modeling of the molecule(s) in the unit cell can be constructed and refined. A unit cell can house one or more molecule(s) depending on how the crystal is packed and it is the smallest repeating unit in a crystal.

3.5.2 Protein crystallization

A key challenge in macromolecule crystallography is to crystallize the subject of study, which may be proteins, nucleic acids or protein-ligand complexes [42]. In order to drive macromolecules to form crystals, a common method used is vapor diffusion. In this method, a small drop of highly concentrated soluble and pure protein solution is mixed with a precipitant in a specific ratio, placed next to a volume of precipitant reservoir in a sealed environment. The difference of vapor pressure between the droplet and precipitant reservoir will gradually drive the droplet concentration to achieve equilibrium in the closed environment. If the condition is optimal, this allows the protein solution to enter supersaturation phase and subsequently undergoes nucleation that will lead to crystal formation [42].

There are many variables that could affect crystal formation [43]. Firstly, the quality of protein is very important. Proteins that are soluble, homogenous and stable (above crystallization temperature) have a higher propensity to form crystals. Other factors to be considered include temperature, pH, precipitant choice, precipitant concentration, precipitant-protein ratio, presence of additive or ligands starting protein concentration, etc. The choice of precipitant and buffer composition is usually made by trial and error. Nowadays, commercially available crystallization screens are used as an initial setup of crystallization trials. Once the initial crystallization conditions are identified, subsequent optimization by altering the variables described above can further improve the quality of the crystals.

3.5.2.1 Crystallization of L-PGDS

L-PGDS was crystallized in the presence 2 mM substrate analog 9,11-epoxymethano PGH_2 (SA-U44069) but no crystals were formed in the absence of SA-U44069 (Figure 8A). In order to study the structural differences of L-PGDS with and without SA-U44069, an apoenzyme structure was needed. Microcrystals of L-PGDS and SA-U44069 complex (SAC)

were streaked on protein-precipitant drop without substrate analog. This is known as seeding and commonly used in protein crystallization to provide nucleation sites for molecules to assemble under the right condition. Crystal grown from the seed usually adopts the characteristics of the template crystals.

3.5.2.2 Crystallization of SMPDL3a

SMPDL3a, with 21.6 mg/ml concentration, took 21 days to form crystals at 24°C. Initial crystals were small and diffracted to 3-3.5 Å resolution (Figure 8B). The optimizations of crystallization condition include changing the additive concentration and temperature of incubation helped to reduce the period of crystal formation down to 7 days and improved the diffraction to 2.3 Å resolution (Figure 8C). This was sufficient for heavy atom based phasing experiments that were used to solve the structure.

3.5.3 Data collection

Once a crystal is obtained of a macromolecule, it is subjected to a focused monochromatic X-ray beam and if it diffracts, the diffraction data can be detected and collected. The basic setup used nowadays has the crystal mounted in a drop containing a cryo-protectant like glycerol, sitting in a cryo loop and frozen under a stream of liquid nitrogen gas. The maintenance of the crystal at cryo temperatures often dramatically increases crystal lift time and data quality. An X-ray beam from a copper source is directed onto the crystal loop mounted on a goniometer that can be rotated in almost any orientation with respect to the beam. A detector then collects the diffraction spots from the rotating crystal (Figure 8D). The intensities of the spots, also known as reflections, can now be integrated to generate a diffraction dataset.

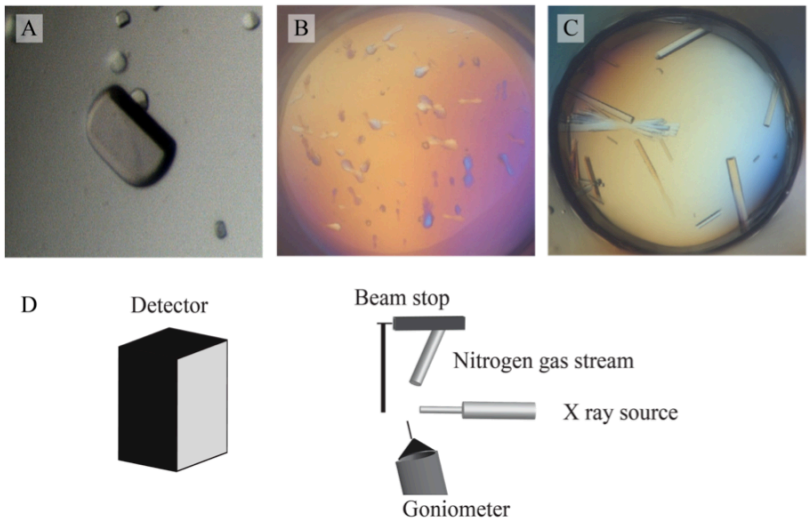


Figure 8. (A) L-PGDS crystal after optimization (B) SMPDL3a crystals without additive (C) SMPDL3a crystal with 0.1M TCEP as additive (D) Basic setup of a synchrotron beamline where the sample is mounted on a goniometer cooled with nitrogen gas stream. When subjected to an X-ray beam, the diffraction pattern is collected on the detector.

X rays can be generated by X-ray tubes, rotating anodes or a synchrotron radiation sources. The synchrotron radiation is where the most powerful and highest-quality X-ray source available. Complete datasets can be collected within minutes at high-end synchrotron stations. Besides, it can produce monochromatic X-ray beam at selected wavelength that is useful for experimental phasing (*Section 3.5.4.2*).

Dataset collections for L-PGDS and SMPDL3a have been carried out at the following synchrotron facilities; Australia Synchrotron (Melbourne, Australia), BESSY Helmholtz-Zentrum (Berlin, Germany), Diamond Light Source (Oxford, UK), European Synchrotron Radiation Facility (Grenoble, France), Soleil (Gif sur Yvette, France) and National Synchrotron Research Radiation Center (Hsinchu, Taiwan).

3.5.4 Solving the phase problem

In order to solve the macromolecule structure, both reflection intensities and the phase information are required. This is depicted in the Fourier summation that generates the electron density in which each reflection from the diffracting crystal is represented by an amplitude and a relative phase (α) (*Section 3.5.1*). The phase information unlike the intensities (amplitudes) cannot be obtained directly using current instruments [44], this lack of phase information is commonly known as the phase problem.

Nonetheless, several methods can be used to solve this problem. Phases can be determined experimentally by isomorphous replacement (single isomorphous replacement, SIR; multiple isomorphous replacement, MIR), anomalous scattering (single wavelength anomalous dispersion, SAD; multi-wavelength anomalous dispersion, MAD) or combination of both, SIRAS / MIRAS. On the other hand, if a related protein structure has been previously solved, molecular replacement can be employed. This section will discuss two specific methodologies in more detail.

3.5.4.1 Molecular Replacement (Paper I)

Molecular replacement (MR) is a method that uses the known phase information of a related structure to generate an initial phase estimate for crystals of a new protein. As a rule of thumb, it is good to have at least 25% of amino acid sequence identity between the previously solved structure - a “search model” and the unknown structure known as a “target molecule” for better chance of success [45]. Normally, the best search model shares more than 30% of sequence identity between the two. MR sequentially employs rotation and translation functions in finding the appropriate orientation and position of the search model in the asymmetric unit (Figure 9). If successful, approximate phases can be calculated from this model and when combined with the experimentally derived structure factor amplitudes of the target molecule’s crystals, an electron density map can be calculated for the target molecule. Subsequently, a model of the target molecule can be generated in this electron density, potential guided by the search model. Finally, this model is improved and refined in a cyclic process. Since mouse L-PGDS that shares 72.5% sequence identity with human L-PGDS had

been previously determined, it made a good search model to solve the structure of human L-PGDS using molecular replacement.

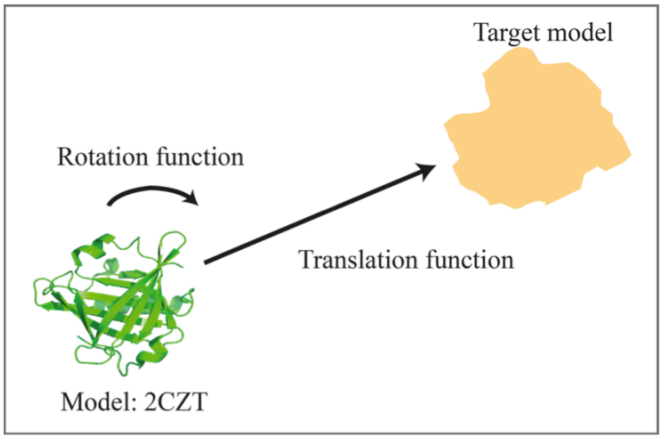


Figure 9. Illustration of rotation and translation functions applied onto mouse L-PGDS (PDB i.d.: 2CZT) as search model to solve the phase of target model, human L-PGDS.

3.5.4.2 Single Anomalous Dispersion (SAD) (Paper II)

Heavy atoms are rich in electrons, and contribute strong anomalous scattering when excited at a wavelength close to the atom's absorption edge. Radiation is scattered with an altered phase due to the anomalous scattering and this altered phase is dependent on the heavy atom position in the unit cell. The site(s) of heavy atoms in the unit cell have to be determined to allow the phases to be calculated [46]. This method requires a good incorporation of heavy atoms into the crystal sample and an X-ray source with a tunable wavelength. One way of adding heavy atoms directly to the protein during protein expression is by using selenium modified methionine amino acids. Alternatively, heavy atoms can be incorporated into the protein crystals through co-crystallization or soaking. Some common choices of heavy atoms include Pt^{2+}/Pt^{4+}, Hg^{2+}, Au^{3+} and Pb^{2+}, which have strong affinity towards amino acids like histidines, cysteines and methionine. They could be added as their salts but also in the form of metalo-organic compounds. Apart from these commonly used heavy atoms, some natural occurring metals such as manganese, copper, iron and zinc are also suitable anomalous scatterers.

SAD phasing requires data collection at the wavelength where absorption peak of the heavy atom is significant (*f'*) and anomalous differences is strong. In MAD phasing, additional diffraction data at the inflection wavelength (*f''*) and remote wavelength is included to maximize the dispersive difference thereby increasing the quality of the phase information. Ideally, all anomalous dataset collection should come from a single crystal to minimize non-isomorphous difference. However, in practice radiation damage usually limits data quality.

Zn^{2+} is found to stabilize the SMPDL3a protein significantly in the thermal shift assay. As Zn^{2+} has a strong anomalous signal (f'' = 3.9 electrons at its *K* edge), it was a primary choice for experimental phasing of SMPDL3a. Twelve heavy atom sites were located in one asymmetric unit cell based on data collected at the Zn edge (9.661 keV). Phases determined

by autoSHARP version 2.60 [47] yielded an electron density map that allows poly-alanine chain building in Coot [48] and subsequently a complete model construction by ARP/wARP version 7.5 from the poly-alanine model (Figure 10) [49].

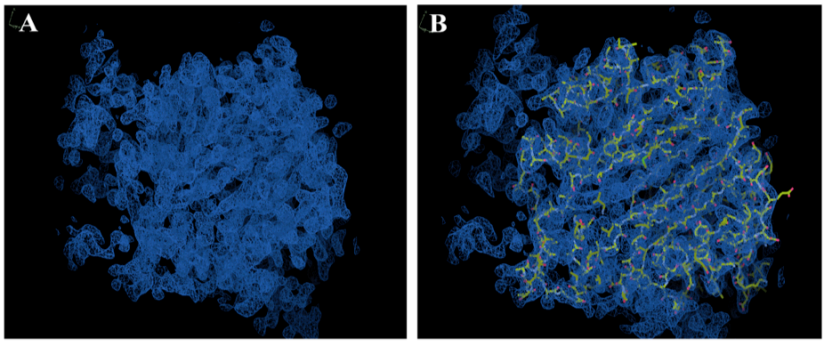


Figure 10. (A) Electron density map from SHARP. (B) Poly-alanine model built in Coot using the map as a guide.

3.6 SOLUTION NMR SPECTROSCOPY

NMR spectroscopy is a powerful and sensitive tool for structural study and is extremely advantageous to complement work of X-ray crystallography. Solution NMR study the protein structure in the solution, hence no crystal is needed. NMR can also provide both qualitative and quantitative measurements of protein-protein or protein-ligand interactions as well as to investigate internal protein dynamics.

Atoms are made of neutron, proton and electron each possessing the property of spin that comes in a magnitude of ½ . Only nuclei with a total non-zero spin are "visible" in NMR spectroscopy, such as ^{1}H, ^{2}H, ^{13}C, ^{15}N and ^{31}P. When these nuclei are placed in the presence of static magnetic field (B_0), the nuclei spins will act as magnetic moment and align in the magnetic field. Under thermodynamics equilibrium, the nuclei spins are populated in high and low energy levels based on a Bolztmann distribution. The energy required for the transition between two levels is represented as follow, where h is Planck constant and γ is the gyromagnetic ratio of nuclei.

$$E = h\nu B_0$$

When an energy that matches the energy difference (E) is applied in a form of radio frequency (RF), the nucleus absorbs the energy into a higher energy state and precesses opposing or perpendicular to B_0 depending on the type of RF irradiation. The precession generates electric current in the detection coil and since magnetization will decay in time to its equilibrium state, the signal is also known as free induction decay (FID). FID is subjected to Fourier transform to produce the NMR spectrum for further analysis [50].

The NMR spectrum comprises of NMR lines located at specific frequency known as chemical shifts, δ. The frequency detected is a "shift" of nucleus magnetization signal from

B_0 due to electron shielding of the molecule's the chemical environment relative to the standard nucleus magnetization. The standard used in L-PGDS experiments is 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Since δ is dependent on the chemical environment of atoms, it is able to reflect the chemical structure of molecules studied. Usually a macromolecules solution NMR experiment would begin with a one-dimensional (1D) experiment followed by two-dimensional (2D) and subsequently three-dimensional (3D) experiments for backbone assignment and structural calculation.

3.6.1 1D experiment : ^{1}H

Since ^{1}H is present in all biological samples, naturally and abundantly, no isotope labeling is required in preparing the sample. Observation of ^{1}H magnetization of the sample is known as 1D experiment. However the high abundance of proton signal from buffer and solvent in the sample dwarfs the protein peak. Most importantly the NMR lines corresponding to protein are highly overlapped and difficult to be interpreted. Therefore it is common to adopt multi-dimensional (2D and 3D) experiments involving ^{1}H , ^{15}N and ^{13}C isotope labeled protein to selectively increase the signal sensitivity and reduce signal overlap from the protein sample. Nonetheless, 1D-experiment is routinely used to enable a brief qualitative overview of the protein secondary and tertiary structure. 1D-NMR spectra of L-PGDS showed a good spread of side chains nitrogen bound protons (H^N) and backbone H^N chemical shifts from 5.5 ppm to 10 ppm indicating the protein is globular and consists of mainly β-sheets structure (Figure 11).

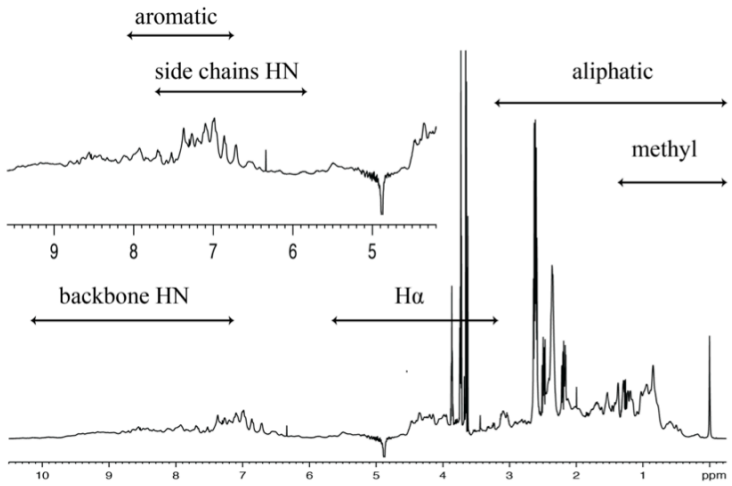


Figure 11: 1D-NMR spectra of 0.38 mM L-PGDS in 50 mM sodium phosphate pH 6.5, 20 mM NaCl, 2mM TCEP , 5% D_2O and 0.5mM DSS

3.6.2 2D experiment: Heteronuclear Single Quantum Coherence (HSQC)

Multidimensional experiments are key to circumvent the limitations of 1D-experiments. They resolve overlapping NMR lines and provide further information of connectivity between the same or different nuclei in the molecule. They serve as a basis for chemical shifts assignment and structural elucidation. Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence

(HSQC) experiment involves the transfer of magnetization from ^{1}H to its directly bonded ^{15}N nuclei and back to ^{1}H for detection (Figure 13B). ^{1}H-^{15}N HSQC spectrum maps the chemical shifts correlation between ^{1}H (x-axis) and its directly bonded ^{15}N (y-axis) in distribution of cross peaks. Since proteins are made of amino acids linked with peptide bonds, the majority of these cross peaks are related to the protein amide backbone. In addition, signals from the side chain NH bonds of tryptophan, arginine, lysine, histidine, glutamine and asparagine are also "visible". However, proline is "invisible" due to its cyclic backbone. Most importantly the positions of these cross peaks are unique to every protein like a fingerprint, because it is based on the amino acids' chemical and structural environment. Therefore it serves as a reference map for chemical shifts assignments in combination with data from other multi-dimensional experiments. Other use of HSQC experiment involves protein quality inspection and protein-ligand interaction studies.

3.6.2.1 Conformational analysis

^{15}N labeled L-PGDS is prepared in buffer 20 mM HEPES pH 6.5, 150 mM NaCl, 2mM TCEP, 10% D_2O and 0.5 mM DSS for ^{1}H-^{15}N HSQC measurement. As mentioned previously in *Section 3.1.3.2*, L-PGDS elution from SEC has a yellow fraction (Fraction 1) and a colorless fraction (Fraction 2) . Even though both fractions appeared as single band in SDS PAGE analysis, their ^{1}H-^{15}N HSQC NMR spectra are distinctively different (Figure 12). The cross peaks from Fraction 2 are better resolved and higher in quantity as compared to Fraction 1. It is likely that these two fractions are conformational heterogeneous. Therefore Fraction 2 was pooled for both NMR and crystallography studies.

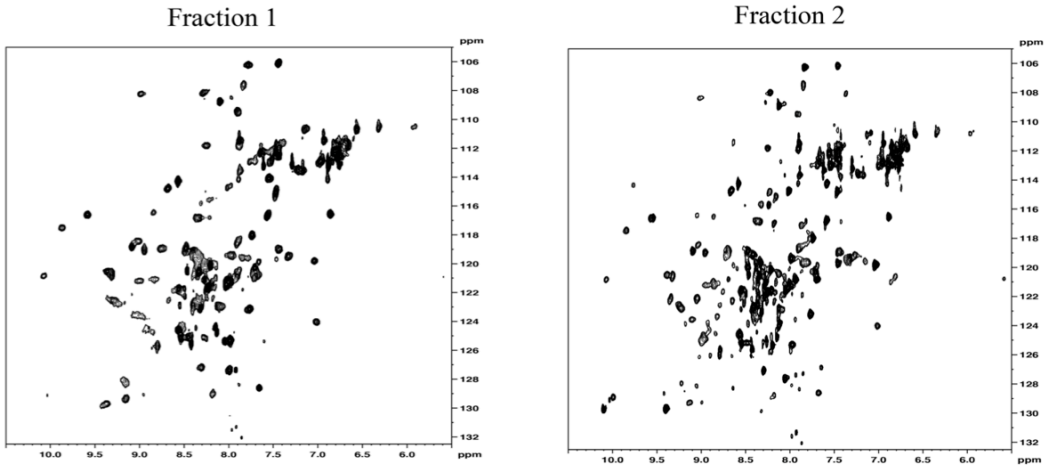


Figure 12: 2D-HSQC spectrum of L-PGDS Fraction 1 and Fraction 2

3.6.2.2 Ligand titration

In the event of ligand addition, the distribution of cross peaks will alter, as chemical shifts of the nuclei are sensitive to structural changes. This is known as ligand-induced chemical shifts perturbation (CSP). The ^{1}H-^{15}N HSQC experiment is able to locate ligand-binding site(s) in the protein of interest if its chemical shifts has been assigned. ^{1}H -^{15}N HSQC experiments of

^{15}N L-PGDS were recorded in the absence and presence of substrate analog (SA U44069) and product analog (PA 12415). Upon the addition of ligands, cross peaks of amino acids involved in binding and conformational changes can be detected (Figure 13A). Both analogs were titrated in concentrations of 2 mM, 3 mM and 4 mM. The CSP was measured for each cross peak and a threshold of 0.1 ppm was marked to distinguish between real binding and background perturbation [51]. It was also observed that HSQC spectrum of ^{15}N L-PGDS with 1 µL of substrate analog has significantly more cross peaks that are more resolved as compared to those without ligand (Appendix I & II). Therefore this spectrum is used as a reference sample for backbone assignment in order to assign as many residues as possible (Appendix III).

Furthermore, in order to test the hypothesis of possible interactions between L-PGDS and membrane during hydrophobic substrate binding and product release, titration of the analog bound ^{15}N L-PGDS complex with 3 mM dodecylphosphocholine (DPC) was carried out. DPC is a common detergent used as a membrane mimetics in NMR, especially in the study of interfacial membrane proteins [52]. It spontaneously forms micelle above its critical micelle concentration (c.m.c = 1.2 mM). Interaction of DPC micelles with ^{15}N L-PGDS was observed by the changes in CSP and alteration of intensity for all detectable cross peaks. Apo-enzyme titration with DPC micelles was used as a control experiment. In order to specifically locate the site of ligand binding and membrane mimetic interactions, sequential assignment of L-PGDS amide backbone was required. The backbone assignment was accomplished by several 3D experiments that will be discussed in the next section.

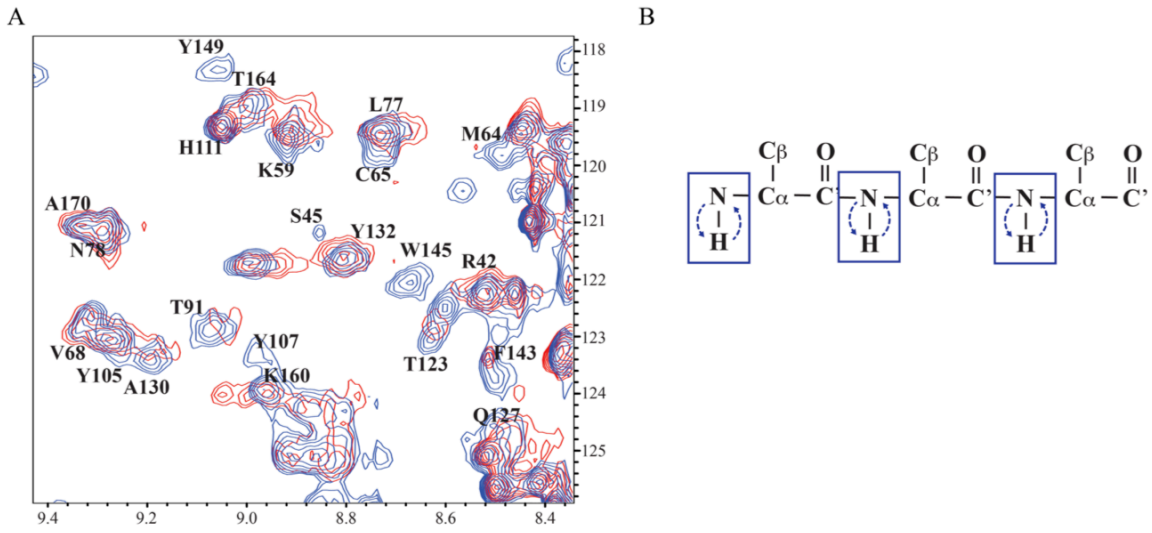


Figure 13: (A) 15N – HSQC of L-PGDS apoenzyme in red and with 2 mM substrate analog U44069 in blue. Residues Y149, M64 and Y107 are among the chemical shifts that showed significant perturbation in the presence of substrate analog. (B) The transfer of magnetization from ^{1}H to its directly bonded ^{15}N nuclei and back to ^{1}H for detection in HSQC experiment.

3.6.3 3D experiments: HNCA, CBCA(CO)NH, ^{15}N NOESY-HSQC, HNCO

3.6.3.1 HNCA and CBCA(CO)NH

Both HNCA and CBCA(CO)NH experiments require ^{13}C and ^{15}N enriched protein sample. The HNCA experiment transfers magnetization via J-coupling from amide proton to amide nitrogen and then to its own alpha carbon, C^{α}_{i} and previous residue C^{α}_{i-1} subsequently back to amide proton for detection (Figure 14B). The magnetization transfer correlates to the amide cross peaks in ^{1}H-^{15}N dimension with its intra and inter-residue Cα in the ^{15}N -^{13}C dimension. The C^{α}_{i} , and C^{α}_{i-1} cross peaks are viewed in strips where the signal of C^{α}_{i} is usually stronger than C^{α}_{i-1}, chemical shifts of these Cα would overlapped if they are adjacent residues (Figure 14A).

Meanwhile the CBCA(CO)NH experiment correlates the amide cross peak with its previous Cα and Cβ residue also via J-couple magnetization transferred (Figure 14B). Together with HNCA, they provide sequential information of the amide cross peaks. Since the chemical shifts of Cα and Cβ are distinguishable between amino acids due to their chemical structure, the identity of amino acids and its sequence can now be pieced together.

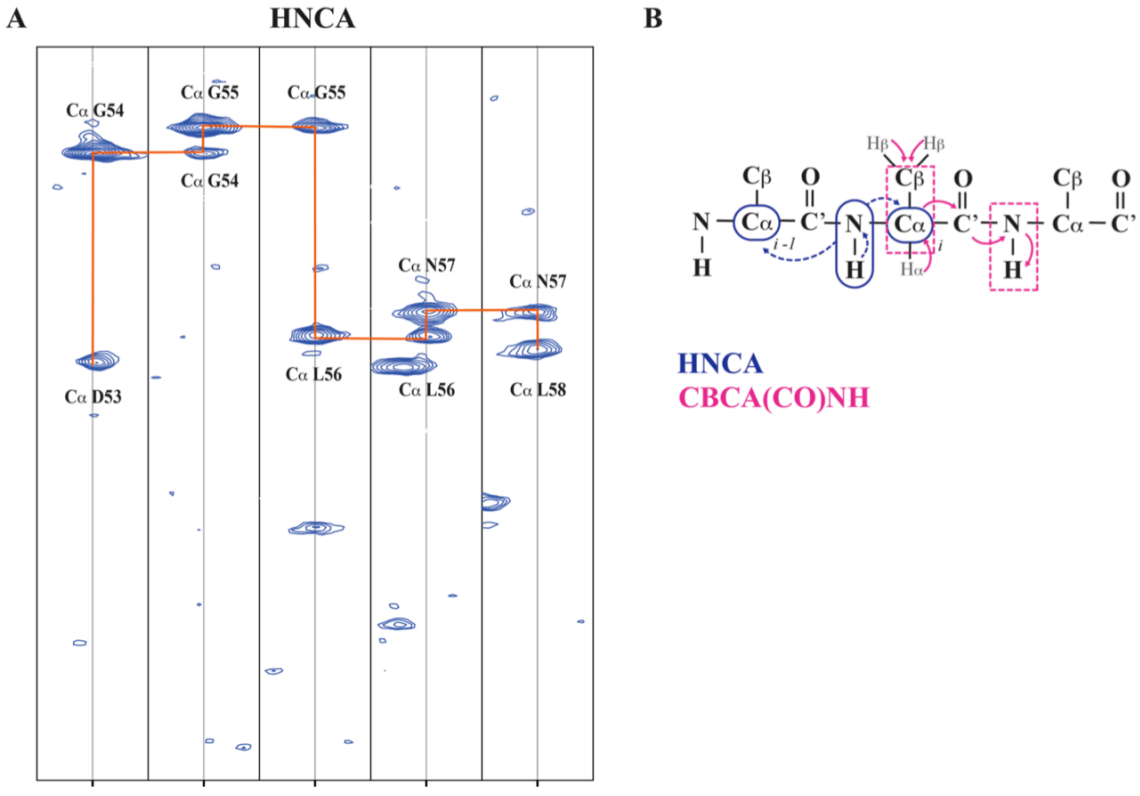


Figure 14: (A) An assigned and connected strip of HNCA in ^{15}N-^{13}C dimension. (B) The magnetization transfer route for both HNCA and CBCA(CO) NH experiments.

3.6.3.2 ^{15}N (NOESY)- HSQC

Nuclear Overhauser effect (NOE) is a phenomenon observed when NMR signal intensity of a nucleus is enhanced due to the effect of resonance frequency saturation of another nucleus in close spatial proximity. This effect provides information of intermolecular distances with the intensity of NOE proportional to $1/r^6$ where r is the distance between two nuclei (usually protons). Therefore, a proton-proton distance within 5Å will give a NOESY signal. A NOESY-HSQC experiment allows magnetization exchange between all protons via NOE and then transferred back to the amide proton for detection. In this case it acquires knowledge of structural proximity in relation to the amide proton. Therefore, together with L-PGDS crystal structure, NOESY conformational dependent information can be used to validate backbone assignment of the protein.

3.6.3.3 HNCO with specific residue labeling

Due to the close Cα Cβ chemical shifts of residues like Leucine and Alanine, a labeling by residue type (LBRT) strategy was adopted to validate their assignments. This method required preparation of protein samples with ^{14}N ^{13}C-carbonyl labeled Leucine (or Alanine in a separate protein expression batch) in ^{15}N M9 minimal media. HNCO experiments transfers magnetization from the amide proton to the preceding carbonyl C (CO). The signals select the ^{15}N labeled residue that comes after ^{14}N ^{13}C-carbonyl labeled Leucine or Alanine in the sequence (Appendix IV & V). This information helps to resolve the ambiguity of assignment between these chemically alike residues.

4 PAPER I: THE STUDY OF LIPOCALIN PROSTGLANDIN D SYNTHASE (L-PGDS)

4.1 L-PGDS IN EICOSANOIDS SIGNALLING

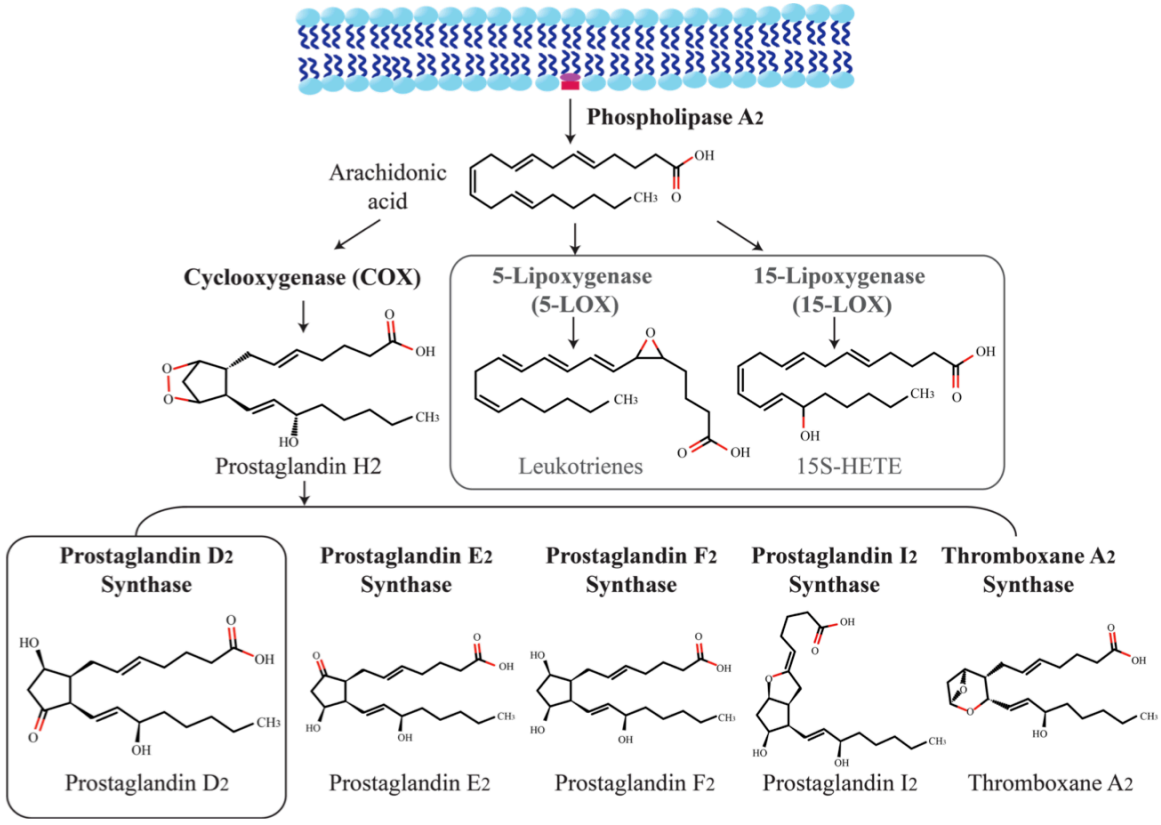


Figure 15: The arachidonic acid pathway.

Eicosanoids signaling is a major lipid-signaling pathway involved in acute inflammatory responses, nociception, platelet aggregation and immunoregulation [53]. External signal like stress triggers cytoplasmic phospholipase to activate membrane phosphatidylcholines lipid into arachidonic acid (AA) [54]. AA can be converted to prostaglandins by cyclooxygenase (COX), or turned into leukotrienes and lipoxins by the action of lipoxygenases (LOX). The product of COX, Prostaglandin H_2 is a common substrate for thromboxane synthase and tissue-specific prostaglandin isomerases like Prostaglandin D_2 synthase, Prostglandin E_2 synthase, Prostaglandin F_2 synthase and Prostaglandin I_2 synthase (Figure 15). Prostaglandins function in autocrine and paracrine manner interacting with cognate prostaglandin receptors on plasma membrane. These ligands controlled G-protein coupled receptor then activates downstream signaling either by intracellular changes in cyclic AMP (cAMP) concentration or Ca^{2+} mobilization [55].

The work in Paper I focused on Lipocalin prostaglandin D_2 synthase (L-PGDS). Two types of PGD_2 synthase (PGDS) have been characterized based on its tissue distribution, hematopoietic PGDS (H-PGDS) and L-PGDS. H-PGDS is highly expressed in blood, mast cells and immune cells while lipocalin PGDS (L-PGDS) is mainly found in the central

nervous system, heart and reproductive organs [56]. They are distinct in overall structure, catalytic site and co-enzyme dependence despite catalyzing the same chemical reaction. H-PGDS catalysis is glutathione dependent but L-PGDS catalysis involves a cysteine residue. L-PGDS also moonlights as lipophilic ligand transporter, binding molecules like retinoic acid, bilirubin and biliverdin. The enzyme catalytic mechanism of L-PGDS has been previously speculated based on mouse apoenzyme structure but the substrate positioning and product delivery is unknown. Paper I addresses substrate binding and mechanistic aspects of L-PGDS substrate binding, catalysis and egression.

4.2 RESULTS SUMMARY

L-PGDS – substrate analog complex (SAC) crystal structures showed two sites of plausible ligand interaction but the density for substrate analog SA U44069 is ambiguous even at 1.8 Å. This uncertainty is confirmed by NMR titration studies with SA U44069 where the interaction sites coincide with what have been observed in the crystal structure. The hydrophobic moieties of the prostanoid substrate (PGH_2) has only three sites for hydrogen bonding inside the substrate-binding pocket; a cyclic endoperoxide group, a hydroxyl group on aliphatic ω chain and a carboxylate group on the acid tail. L-PGDS has a large hydrophobic beta barrel pocket relative to the size of the substrate. This increases the degree of freedom for PGH_2 and SA flexible hydrophobic tails to reside in the binding pocket. Furthermore, SA U44069 comprises of a non-hydrolysable bicyclicpentane ring that reduces it's the number of hydrophilic anchoring sites. It is possible that all these factors contribute to the poor ligand occupancy and discontinuous difference density (Fo-Fc) of SA U44069 in the crystal structure. Modeling of SA U44069 in the active sites identified three key areas for interactions. First, the bicyclopentane head group can be stabilized by Phe 83 to be positioned facing Cys 65 for catalysis. Secondly its carboxylate tail can be anchored by polar residue like Tyr 107 or His 111 and lastly the aliphatic hydrophobic chain can be inserted deep into the barrel lined with serine and leucine residues.

Co-crystal structures of L-PGDS also revealed important conformational changes of the Ω helix when compared to the apoenzyme structure. This observation was also supported by NMR titration results with both SA U44069 and product analog (PA 12415) whereby cross peaks of residues on Ω helix undergo CSP due to ligand-induced binding. Furthermore, CSP analysis also supported the "stop-plug" conformational changes proposed by Zhou et al where the residues at the bottom of the beta barrel undergo measurable structural changes in the event of ligand binding [57]. Similar patterns of ligand-induced CSP were observed when the L-PGDS-PA complex was titrated with the DPC micelles. In fact, upon DPC micelles titration, CSP of L-PGDS-PA complex were more pronounced than those of L-PGDS-SA complex. The peak intensity analysis for DPC-L-PGDS-PA spectrum showed that for some residues, the intensity enhancement of cross peaks during initial addition of PA to an apoenzyme were reversed or further attenuated when DPC micelles were added to the PA-enzyme complex. It is likely due to the reversal of PA binding upon DPC micelles titration. The contrasting CSP and cross peaks intensity alteration between DPC-L-PGDS-PA and

DPC-L-PGDS-SA experiments revealed different enzyme-ligand and enzyme-membrane interaction depending on the type of cargo bound.

Enzymes are known to have developed diverse strategies in handling substrate and product of seemingly opposing chemical properties [58]. This cargo-dependent interaction may likely reflect the *in-vivo* transport nature of soluble L-PGDS utilized for hydrophobic enzymatic product delivery. After catalysis, the enzyme may release its product PGD_2 into endoplasmic reticulum membrane with transient attachment facilitated by Trp 54 on Ω loop and Trp 112 on E-F loop, and potentially undergo conformational changes on Ω helix and E-F loop. PGD_2 may be delivered to plasma membrane via vesicular transport and released to the extracellular space by multi-drug resistant transporters [59]. This would allow PGD_2 to activate downstream signaling through cell surface receptors DP1 or CRTH2 (Figure 16) [60, 61]. All in all, an integrative structural biology strategy had provided a more holistic and resolved view of L-PGDS function, both as an enzyme and a transporter.

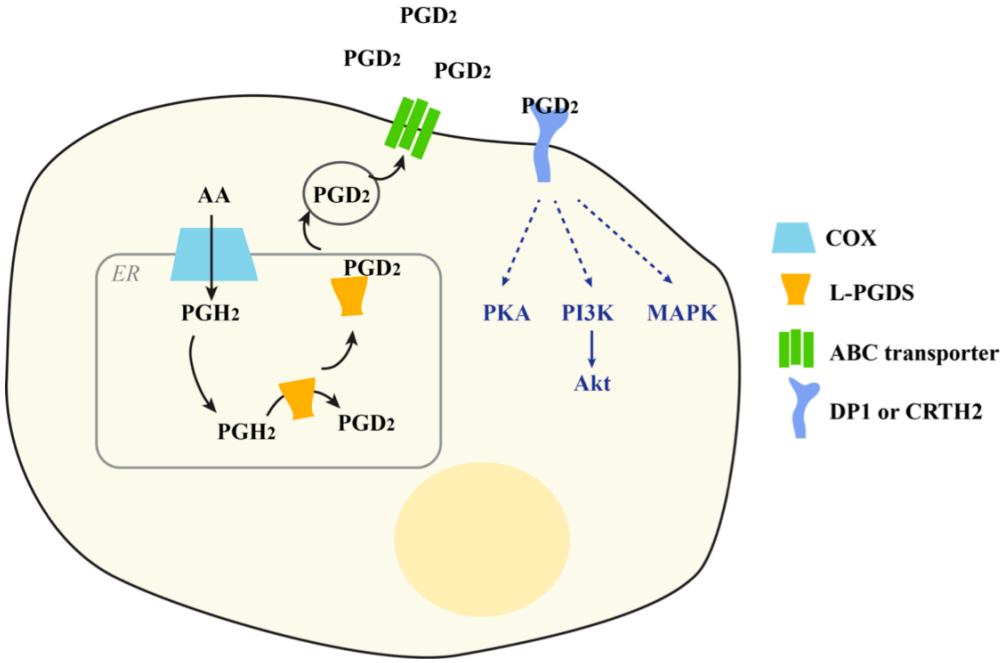


Figure 16. Proposed model of L-PGDS as an enzyme and transporter in the cell.

5 PAPER II: UNRAVELING A NOVEL PHOSPHODIESTERASE IN STRUCTURE & FUNCTION

5.1 SPHINGOMYELINASE PHOSPHODIESTERASE LIKE 3A (SMPDL3A)

Since the published of human genome in 2001, transcriptomic and proteomic studies have been accelerated by the wealth of genomic information [62, 63]. Collaborative effort in bridging the genomics proteomics gap has been ongoing for the past fourteen years. Presently, there are many proteins which their function and regulation is still unknown. Computational driven function prediction for uncharacterized protein have gone beyond sequence homology to include cell distribution, phylogenetic profiles, protein's domain analysis and gene position in chromosomes to provide a framework for experimentalist to plan their investigation [64]. Unknown enzymes can infer its function based on such predictions and are further grouped in respective family. Sphingomyelin phosphodiesterase like 3a (SMPDL3a) belongs to the aforementioned. With only a few published papers in the literature, it has been associated based on sequence homology and protein domain conservation to the function of the well-studied sphingomyelin phosphodiesterase 1 (SMPD1) or commonly known as acid sphingomyelinase (aSMase).

5.1.1 ASMase in sphingolipid hydrolysis

ASMase is an enzyme in the sphingolipid pathway that cleaves the phosphodiester bond on sphingomyelin to yield ceramide, an important signaling sphingolipid. Ceramide can activate cell apoptosis, senescence or cell-cycle arrest. It can be converted further to ceramide-1-phosphate, sphingosine or sphingosine-1-phosphate to regulate diverse cellular responses. ASMase together with other sphingolipid transforming enzymes are extremely important agents in governing the level of ceramide in the cells. Nonetheless, the structural basis for the function of aSMase is still unknown.

Dysfunctional aSMase causes lysosomal accumulation of sphingomyelin, resulting in a metabolic disorder known as Niemann Pick disease (NPD) type A or type B. Patients suffering from NPD-A and B are found to possess missense mutations on their aSMase gene leading to the loss of function phenotype [65]. In more severe cases, these mutations are detrimental to the neurophysiology of patients and cause death in their early childhood. Unfortunately, the molecular effects of the disease causing mutations is still poorly understood.

5.1.2 Calcineurin-like-phosphodiesterase family

aSMase and SMPDL3a share significant sequence homology (31%) and highly conserved binuclear metal center containing phosphodiesterase domain, leading to the presumption that SMPDL3a operates in a similar manner. The phosphodiesterase domain scaffolds a catalytic dimetal site that composed of primarily histidine, aspartate and glutamate residues for metal

coordination [66]. This domain was first identified in the Ca^{2+} dependent serine/threonine protein phosphatase - calcineurin, therefore also known as calcineurin like phosphodiesterase domain. Despite structural similarities, many enzymes with this conserved domain were found to vary in their substrate selection ranging from phosphorylated proteins, cyclic nucleotides, glycerophosphoesters, sphingolipids, nucleotide phosphates and its metabolites. Furthermore they also have diverse metal cofactor composition including Fe^{2+}, Mn^{2+}, Ni^{2+}, Mg^{2+}, Zn^{2+}, Ca^{2+} and combinations thereof [66]. These differences led to further sub-classification based on substrate and cofactor dependence, having groups like purple acid phosphatase (PAP), sphingomyelinase, nucleotidase, cyclic phosphodiesterase, alkaline phosphatase and many others (Figure 17). In 2014, Traini et al showed that despite sharing the same catalytic domain as predicted from sequence similarity, SMPDL3a has no sphingomyelinase activity [31]. In a very recent study, it is reported that SMPDL3a is the key enzyme in activating a prodrug that targets fructose-1,6-bisphosphophosphate (FBPase) for Type II diabetes treatment [67]. The work in paper II reported the structure of SMPDL3a, as the first structure of this novel sub-class of the calcineurin-like phosphodiesterase. Based on the structural and biochemical information, a possible mechanism that distinguished SMPDL3a from aSMase was proposed.

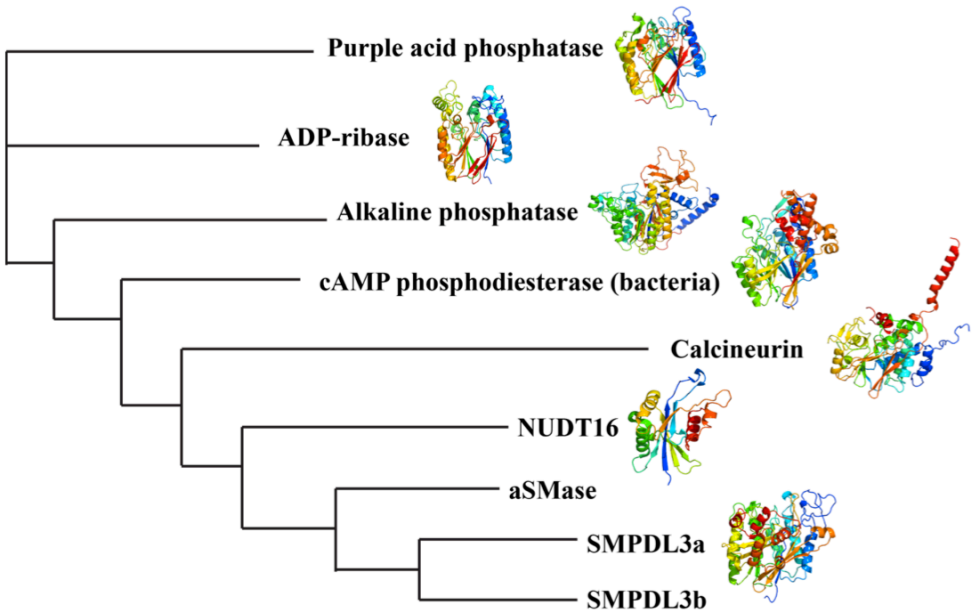


Figure 17. Multiple sequence alignment of SMPDL3a with other members of the calcineurin-like-phosphodiesterase family is represented in a cladogram generated by *Clustal Omega*.

5.2 RESULTS SUMMARY

The crystal structure of human SMPDL3a revealed the expected calcineurin like metallophosphodiesterase fold with two metals coordinated by six histidines, two aspartates and one asparagine. These residues are conserved among the aSMase and aSMase-like proteins, which are distinct as compared to other metallophosphodiesterase proteins. The

protein maintains a βαβαβ motif that comprises almost equal number of α helices and β strands with interspersed flexible loops. There are four N-acetylglucosamines observed in each protein. The functional role of these glycosylation sites is unknown although this glycosylation coat has been shown to protect the protein from thermal induced denaturation. The metals bound in native crystals were identified to be of Zn element based on a fluorescence scan. Additional Zn^{2+} (mM concentration) stabilized the protein significantly but inhibits its catalytic activity.

This study confirmed that the substrate of recombinant SMPDL3a *in vitro* is ATP but not sphingomyelin. Furthermore it is found to be able to hydrolyze substituted nucleotides such as CDP-choline, ADP-ribose and CDP-ethanolamine. These modified nucleotides are known substrates for another calcineurin like member, Mn-dependent ADP-ribose diphosphatase [68]. In order to better understand the substrate binding and catalysis mechanism, the modified nucleotide and various ATP analogs were soaked into SMPDL3a crystals to obtain protein-ligand complexes. In the best complex, CMP moieties, derived from CDP-choline, can be modeled into the ligand density of molecule A and C. Meanwhile only density that corresponds to phosphoribose was observed in molecule B (Figure 18) The CMP product is likely to be trapped inside the protein when the crystal was snap frozen after catalysis. Enzymatic assay and mass spectrometry analysis of the products confirmed that SMPDL3a hydrolysis of CDP-choline yields phosphocholine and CMP. The base moiety of CMP reveals specific interactions with the enzyme in molecule A. Meanwhile the ligands' α-phosphates interact with the active site's metals, His 114 and His 152 residues in all three molecules of the asymmetric unit (Figure 18). Residues His 114 and His 152 are likely to play a key role anchoring the substrate at the metal site for catalysis, they are also conserved in aSMase .

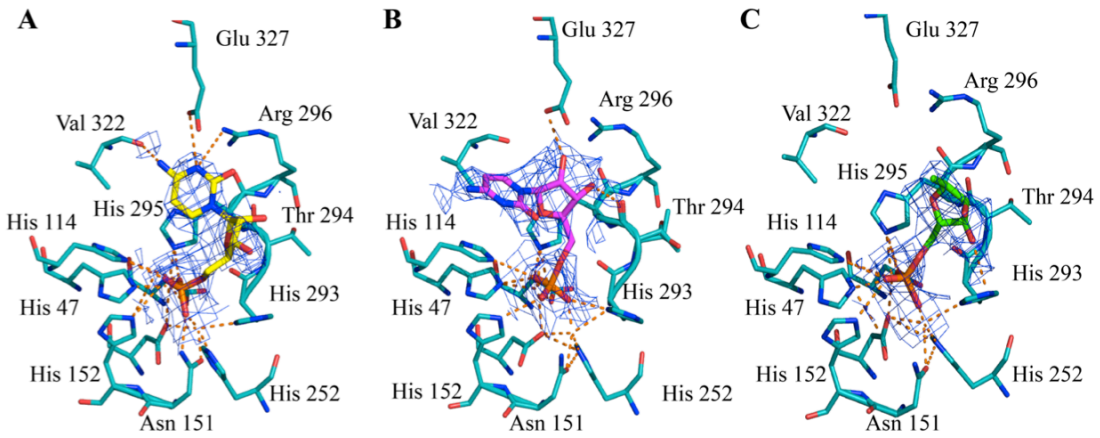


Figure 18. (A) Molecule A of the three subunits of CMP-bound SMPDL3a crystal structure showed interaction of CMP base with enzyme active site residues. (B) CMP in molecule C showed less electrostatic interaction of CMP base and enzyme. (C) Only electron density corresponding to the phosphoribose moiety is observed in molecule B.

The newly available structural data of SMPDL3a can also be used as template for aSMase homology modeling. Strong structural and sequence conservation at the binuclear catalytic

site supports similarities in the chemical mechanism between aSMase and SMPDL3a while large differences in the rest of the substrate-binding site suggest that they are likely to catalyze different substrates. The model also allowed us to map known NPD mutations onto the model of aSMase to shed light on possible molecular basis of the enzyme inactivation in cases whereby mutated proteins are expressed and shuttled to lysozyme like the wild type aSMase [69].

Overall, the structural and biochemical data provide better understanding on the function and catalytic mechanism of SMPDL3a. It is still early to completely rule out the sphingomyelinase activity of SMPDL3a based on current *in vitro* studies. However, other activities, such as hydrolysis of nucleotides, appear more likely based on the present data. Further work is however required for conclusive determination of the *in-vivo* function of SMPDL3a.

6 ACKNOWLEDGEMENTS

My journey of scientific research would not be possible if not for a bunch of wonderful people whom I am forever grateful of. They have filled this journey with many hues and flavors, sharing my ups and downs along the way.

First, I would like to thank my supervisor **Pär Nordlund** for giving me this opportunity to do a joint Ph.D in Karolinska Institutet and NTU, supporting me and grooming me to be an independent researcher. To my co-supervisor **Konstantin Pervushin** who is always full of creative ideas, thank you for trusting me in my work and always readily to discuss and share your NMR expertise with me.

I would like to thank **Hsiangling T** for being such a wonderful mentor in science and in life. The wonderful memories we shared in the lab and in conference will be always in my mind. Many thanks also to **Kit Y** for being such a helpful and hardworking colleague to work with as we crushed the walls of our challenging projects. To **Andreas L**, **Anna J** and **Sue-li D**, I am thankful for your advices and encouragement when I faced difficulties. And thank you **Kelly H**, for sharing my joy and distress throughout my Ph.D and ever so willing to teach, to encourage and to pray for me. I also appreciate **David G** for taking the time to teach me the tips and tricks of experimental phasing on my second project. Many thanks to the **PPP** old and new members, **Tobias C**, **Agnes H**, **Martina N**, **Lucy P**, **Dina D** and **Daniel S** for helping out with the initial screening of my target constructs in Singapore. I will also remember the good times during our lunch sessions in Biopolis with **Stephanie L**, **Dan C** and **Jeslin P**.

As for my colleagues in TROSY lab, **Baiyang**, **Leo W**, **Alistair I**, **Margaret P**, **Shubadra P**, **Rubing L**, **Edward T**, **Li Yan** and **Zhao Jing**, I have learnt so much from your inputs and experiences during our NMR study group and discussions. To **David L**, thank you for guiding me throughout my first few NMR experiments and imparting me with basic scientific skills that I am still holding on. Thank you, **Chandra Verma**, for your encouragement whenever we met up in Biopolis and on your visit to Stockholm.

During the last two and half years of my PhD in Stockholm, I would not be able to swiftly adjust and focus on my research work without many whom have helped me greatly. Thanks to **Lionel T** who is always willingly avail himself for meeting when I need scientific guidance, be it at the corridor or in coffee room. I am also grateful to the responsible **Pelle**, who always maintains the lab order diligently and ever so patient in dealing with all my administrative questions. To the bubbly and cheerful **Fatma G**, a great companion to have in the same office and always generously shares your laughter from KI to all synchrotrons. To **Rebecca D**, our gym trainer and pastry chef, thanks for getting me into the gym to work off some fats before eating your cakes, it's worth it! To **Rozbeh J** and **Michael R**, the men of our Ekeberg fishing team, thanks for all the hardwork and 'patience' in accommodating my 'honesty'. Thanks to **Christian L**, **Jens F**, **Daniel M**, **Esben Q**, **Catrine S** and **Michael D**

for injecting the lab with jokes, gossips, and craziness making 'labbet' such a relax place to work in. To the new members, **Sara**, **Annette**, and **Johan**, good luck to you all and may the new era of PN group be another exciting chapter. To our neighbor Divne group – **Christina D**, **Rosaria G**, **TC** and **Noor**, PN lab would not be complete without your pleasant company. Thanks to **Rosaria G**, my lab "mama", lunch buddy and synchrotron partner for your encouragement and support. To 'uncle' **TC (Tien-Chye)**, whom I can speak Singlish with, your presence makes me feel not too far from Singapore. I am also grateful to the **PSF** members in KI, the reliable **Martin M** for taking care of the crystallization facilities, keeping a conducive environment for all crystallization projects to run smoothly; **Tomas** for assisting me in working with various biophysical machines; **Helena**, **Emma**, **Eleonora** and **Ida** for helping me in construct screening for various targets. To **Herwig**, **Aja**, **Torun**, **Filipa** and **Tobias K,** thank you for being warm and welcoming to me and ever so helpful when I needed to use some instruments from your group.

To my best friends since varsity years, **Kek Yee L**, **Eugene C**, **Kenshaun Y** for sharing my scientist's woes no matter where I am geographically. To my dearest family-in-Christ, **Serene T**, **Yixian H**, **Susan Y**, **Sharmaine Z**, **Paul L**, **Sarah S**, **Kenny L**, **Emma N**, **Wilson W** and **James C**, thanks for sharing the major phase of my doctoral life in Singapore. To my "sisters" (bonded in Vietnam): **Alycia F**, **Xinyi Y** and **Rachel F**, you girls always warm my heart with all the Whatsapp messages and pictures though we are apart.

Thanks to **Lena & Björn D**, **Diane**, **Ellen**, **Shannon** and **Michelle** who warmly welcome me to Stockholm with Singapore food. I am also blessed to have known **Denise L**, **Grace C** and **Yan Han**, whom together we shared the many wonderful gatherings. To **Mariam Z**, the first person I got in touch in Stockholm and have helped me so much in settling down, you are truly an angle. To **Deborah & Pieter T**, **Polly W**, **Nienke K**, **Jessica P**, **Faezzah**, **Nurzian** & **Andreas B**, **Sharon**, **Jingmei**, **Sam & Jinny**, **Xiaohui S**, **Evelyn S**, **Ying Chun**, **Saba**, **Madhan**, **Sandra W**, **Ida H**, **Nina & Mirco**, **Linnea F**, **Joseph T & Cindy**, **Alvin T**, **Ming Wei** and **Yossa**, thank you for blessing me in one way or another in Stockholm.

To my dearest family **Mum**, **Dad** and **Belle** who seen me through this journey, loving me and praying for me, thank you for never ceasing to encourage me and lend a listening ear when I am in distress. Of course to **Jia Chi Y**, thank you for your love, encouragement and support that keep me going strong and inspire me to be the best I can be.

Lastly and most importantly, all glory and praise to **God** for His faithfulness, provision and grace throughout all these years. It would be impossible without Him!

7 REFERENCES

1. de Réaumur, R., *Observations sur la digestion des oiseaux.* Histoire de l'academie royale des sciences, 1752. **1752**(266).
2. Payen A, P.J.F., *Memoir on diastase, the principal products of its reactions, and their applications to the industrial arts.* . Annales de chimie et de physique, 1833. **2**(53): p. 19.
3. Pasteur, L., *Nouveaux faits concernant l'histoire de la fermentation alcoolique.* Annales de Chimie et de Physique, 1858. **3**(52): p. 14.
4. Berzelius, J.J., *Årsberättelsen om framsteg i fysik och kemi.* Royal Swedish Academy of Sciences, 1835.
5. Buchner, E., *Alcholic fermentation without yeast.* Chem. Ber, 1897. **30**: p. 8.
6. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme.* Berichte der Deutschen Chemischen Gesellschaft 1894. **27**(9): p. 2985.
7. Michaelis, L., et al., *The original Michaelis constant: translation of the 1913 Michaelis-Menten paper.* Biochemistry, 2011. **50**(39): p. 8264-9.
8. Sumner, J.B., *The Isolation and Crystallization of the Enzyme Urease.* The Journal of Biological Chemistry, 1926(69): p. 6.
9. Sumner, J.B., *The Digestion and Inactivation of Crystalline Urease by Pepsin and by Papain.* The Journal of Biological Chemistry, 1932(98): p. 9.
10. Northrop, J.H., *Crystalline Pepsin.* Science, 1929. **69**(1796): p. 580.
11. Stanley, W.M., *Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus.* Science, 1935. **81**(2113): p. 644-5.
12. Cech, T.R., *A model for the RNA-catalyzed replication of RNA.* Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(12): p. 4360-3.
13. Bernal, J.D., Crowfoot D. , *X-Ray Photographs of Crystalline Pepsin.* Nature, 1934(133): p. 2.
14. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
15. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-6.
16. Blake, C.C., et al., *Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution.* Nature, 1965. **206**(4986): p. 757-61.
17. Johnson, L.N., Phillips, D.C., *Structure of Some Crystalline Lysozyme-Inhibitor Complexes determined by X-Ray Analysis and 6A Resolution.* Nature, 1965(206): p. 2.
18. Matthews, B.W., et al., *Three-dimensional structure of tosyl-alpha-chymotrypsin.* Nature, 1967. **214**(5089): p. 652-6.
19. Kraut, J., *Serine proteases: structure and mechanism of catalysis.* Annual review of biochemistry, 1977. **46**: p. 331-58.

20. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(2): p. 98-104.

21. Hult, K. and P. Berglund, *Enzyme promiscuity: mechanism and applications.* Trends in biotechnology, 2007. **25**(5): p. 231-8.

22. Reid, T.W. and D. Fahrney, *The pepsin-catalyzed hydrolysis of sulfite esters.* Journal of the American Chemical Society, 1967. **89**(15): p. 3941-3.

23. Nakagawa, Y. and M.L. Bender, *Modification of alpha-chymotrypsin by methyl p-nitrobenzenesulfonate.* Journal of the American Chemical Society, 1969. **91**(6): p. 1566-7.

24. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities.* Chemistry & biology, 1999. **6**(4): p. R91-R105.

25. Copley, S.D., *Enzymes with extra talents: moonlighting functions and catalytic promiscuity.* Current opinion in chemical biology, 2003. **7**(2): p. 265-72.

26. Gamage, N.U., et al., *The structure of human SULT1A1 crystallized with estradiol. An insight into active site plasticity and substrate inhibition with multi-ring substrates.* The Journal of biological chemistry, 2005. **280**(50): p. 41482-6.

27. Theodossis, A., et al., *The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in Sulfolobus solfataricus.* The Journal of biological chemistry, 2004. **279**(42): p. 43886-92.

28. Guillas, I., et al., *Human homologues of LAG1 reconstitute Acyl-CoA-dependent ceramide synthesis in yeast.* The Journal of biological chemistry, 2003. **278**(39): p. 37083-91.

29. Kitatani, K., J. Idkowiak-Baldys, and Y.A. Hannun, *The sphingolipid salvage pathway in ceramide metabolism and signaling.* Cellular signalling, 2008. **20**(6): p. 1010-8.

30. Tomiuk, S., et al., *Cloned mammalian neutral sphingomyelinase: functions in sphingolipid signaling?* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(7): p. 3638-43.

31. Traini, M., et al., *Sphingomyelin phosphodiesterase acid-like 3A (SMPDL3A) is a novel nucleotide phosphodiesterase regulated by cholesterol in human macrophages.* The Journal of biological chemistry, 2014. **289**(47): p. 32895-913.

32. Berman, H.M., et al., *The Protein Data Bank.* Nucleic acids research, 2000. **28**(1): p. 235-42.

33. Savitsky, P., et al., *High-throughput production of human proteins for crystallization: the SGC experience.* Journal of structural biology, 2010. **172**(1): p. 3-13.

34. Jarvis, D.L., *Baculovirus-insect cell expression systems.* Methods in enzymology, 2009. **463**: p. 191-222.

35. Luckow, V.A., et al., *Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli.* Journal of virology, 1993. **67**(8): p. 4566-79.

36. Harrison, R.L. and D.L. Jarvis, *Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins.* Advances in virus research, 2006. **68**: p. 159-91.

37. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Analytical biochemistry, 2006. **357**(2): p. 289-98.

38. Vedadi, M., et al., *Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(43): p. 15835-40.

39. Benjwal, S., et al., *Monitoring protein aggregation during thermal unfolding in circular dichroism experiments.* Protein science : a publication of the Protein Society, 2006. **15**(3): p. 635-9.

40. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions.* Methods in cell biology, 2008. **84**: p. 79-113.

41. Bragg, W.H. and W.L. Bragg, *The Reflection of X-rays by Crystals.* Proceedings of The Royal Society A, 1913. **88**(605): p. 10.

42. Bergfors, T., ed. *Protein Crystallization*. 2009, International University Line: San Diego.

43. McPherson, A., *Introduction to protein crystallization.* Methods, 2004. **34**(3): p. 254-65.

44. D. Sherwood, J.C., *Crystals, X-rays and Proteins.* 2011, Oxford: Oxford University Press.

45. Scapin, G., *Molecular replacement then and now.* Acta crystallographica. Section D, Biological crystallography, 2013. **69**(Pt 11): p. 2266-75.

46. Rhodes, G., *Crystallography made crystal clear: A guide for Users of Macromolecular Models.* 2000, San Diego: Academic Press.

47. Vonrhein, C., et al., *Automated structure solution with autoSHARP.* Methods in molecular biology, 2007. **364**: p. 215-30.

48. Emsley, P., et al., *Features and development of Coot.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 4): p. 486-501.

49. Langer, G., et al., *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.* Nature protocols, 2008. **3**(7): p. 1171-9.

50. Wuthrich, K., *NMR of Protein and Nucleic Acids*1986, USA: John WIley & Sons Inc.

51. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2013(73): p. 16.

52. Warschawski, D.E., et al., *Choosing membrane mimetics for NMR structural studies of transmembrane proteins.* Biochimica et biophysica acta, 2011. **1808**(8): p. 1957-74.

53. Harizi, H., J.B. Corcuff, and N. Gualde, *Arachidonic-acid-derived eicosanoids: roles in biology and immunopathology.* Trends in molecular medicine, 2008. **14**(10): p. 461-9.

54. Wymann, M.P. and R. Schneiter, *Lipid signalling in disease.* Nature reviews. Molecular cell biology, 2008. **9**(2): p. 162-76.

55. Jabbour, H.N. and K.J. Sales, *Prostaglandin receptor signalling and function in human endometrial pathology.* Trends in endocrinology and metabolism: TEM, 2004. **15**(8): p. 398-404.

56. Urade, Y. and N. Eguchi, *Lipocalin-type and hematopoietic prostaglandin D synthases as a novel example of functional convergence.* Prostaglandins & other lipid mediators, 2002. **68-69**: p. 375-82.

57. Zhou, Y., et al., *Structure-function analysis of human l-prostaglandin D synthase bound with fatty acid molecules.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2010. **24**(12): p. 4668-77.

58. Forneris, F. and A. Mattevi, *Enzymes without borders: mobilizing substrates, delivering products.* Science, 2008. **321**(5886): p. 213-6.

59. de Waart, D.R., et al., *Multidrug resistance associated protein 2 mediates transport of prostaglandin E2.* Liver international : official journal of the International Association for the Study of the Liver, 2006. **26**(3): p. 362-8.

60. Matsuoka, T., et al., *Prostaglandin D2 as a mediator of allergic asthma.* Science, 2000. **287**(5460): p. 2013-7.

61. Sawyer, N., et al., *Molecular pharmacology of the human prostaglandin D2 receptor, CRTH2.* British journal of pharmacology, 2002. **137**(8): p. 1163-72.

62. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

63. Tyers, M. and M. Mann, *From genomics to proteomics.* Nature, 2003. **422**(6928): p. 193-7.

64. Eisenberg, D., et al., *Protein function in the post-genomic era.* Nature, 2000. **405**(6788): p. 823-6.

65. Ferlinz, K., R. Hurwitz, and K. Sandhoff, *Molecular basis of acid sphingomyelinase deficiency in a patient with Niemann-Pick disease type A.* Biochemical and biophysical research communications, 1991. **179**(3): p. 1187-91.

66. Wilcox, D.E., *Binuclear Metallohydrolases.* Chemical reviews, 1996. **96**(7): p. 2435-2458.

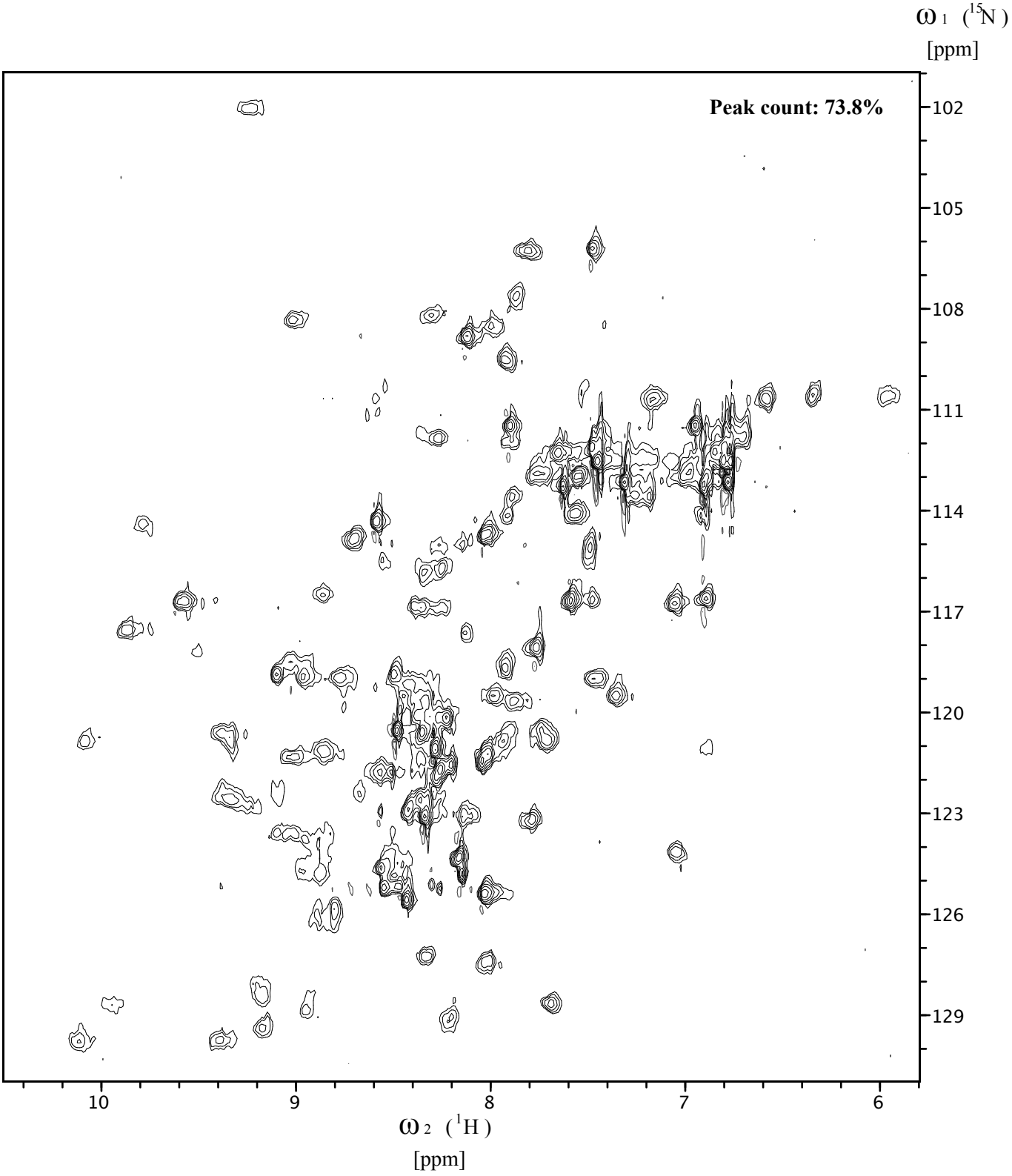
67. Kubota, K., et al., *Identification of activating enzymes of a novel FBPase inhibitor prodrug, CS-917.* Pharmacology research & perspectives, 2015. **3**(3): p. e00138.

68. Canales, J., et al., *Mn2+-dependent ADP-ribose/CDP-alcohol pyrophosphatase: a novel metallophosphoesterase family preferentially expressed in rodent immune cells.* The Biochemical journal, 2008. **413**(1): p. 103-13.

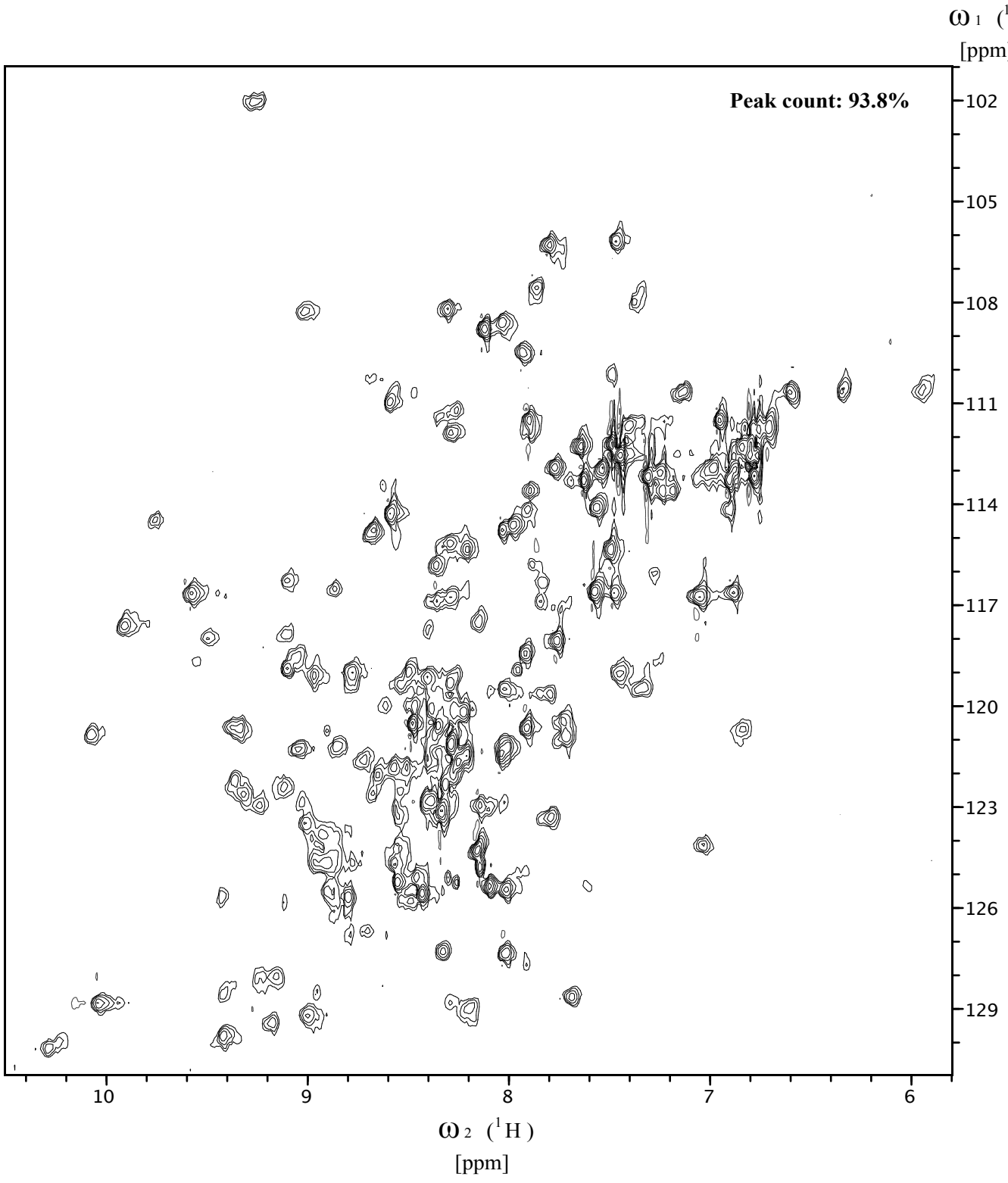
69. Ferlinz, K., et al., *Molecular analysis of the acid sphingomyelinase deficiency in a family with an intermediate form of Niemann-Pick disease.* American journal of human genetics, 1995. **56**(6): p. 1343-9.

8 APPENDICES

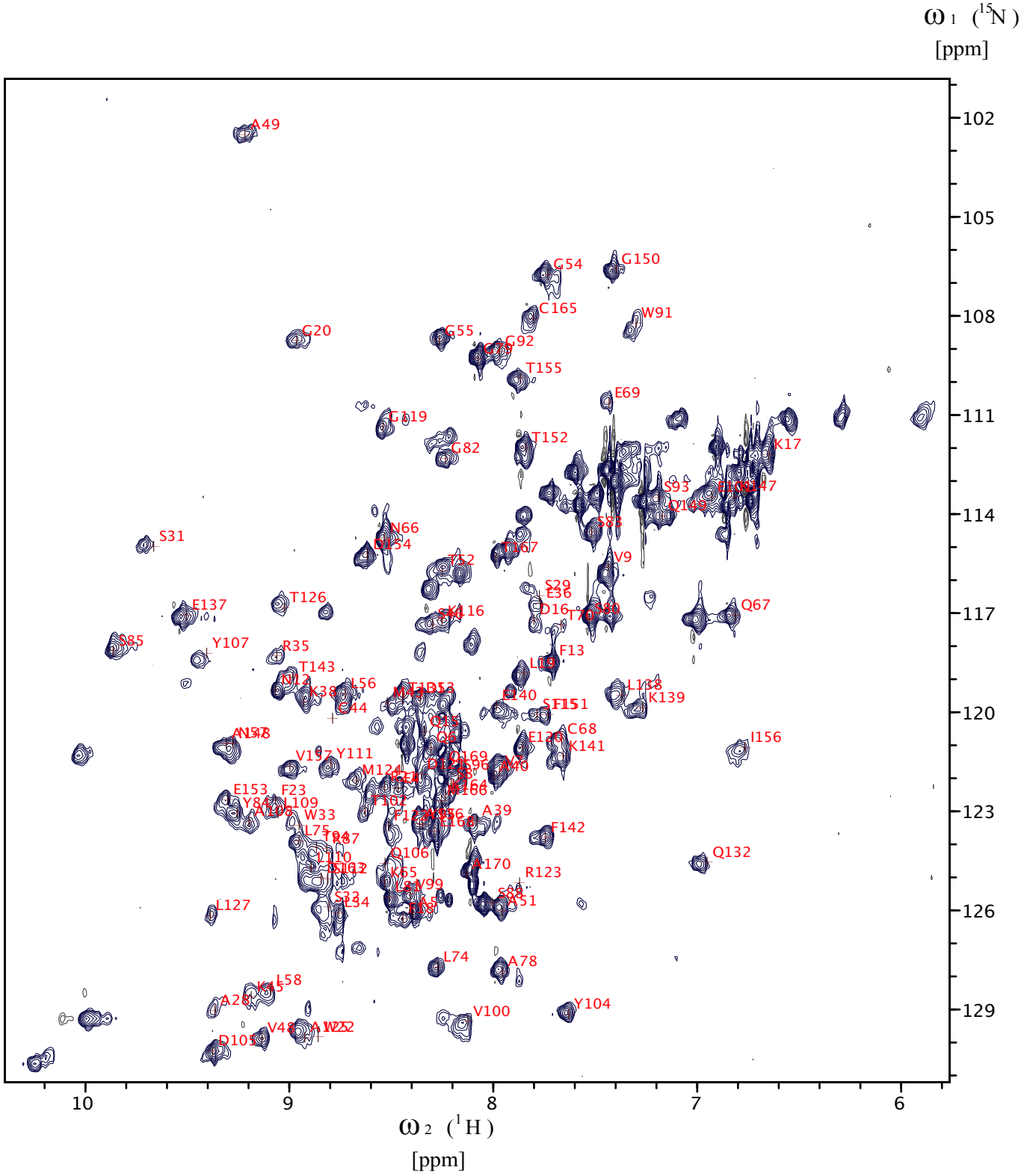
Appendix I : 2D HSQC of 0.39 mM L-PGDS in 20 mM HEPES pH 6.5, 150 mM NaCl, 2 mMTCEP, 10% D_2O, 0.5 mM DSS



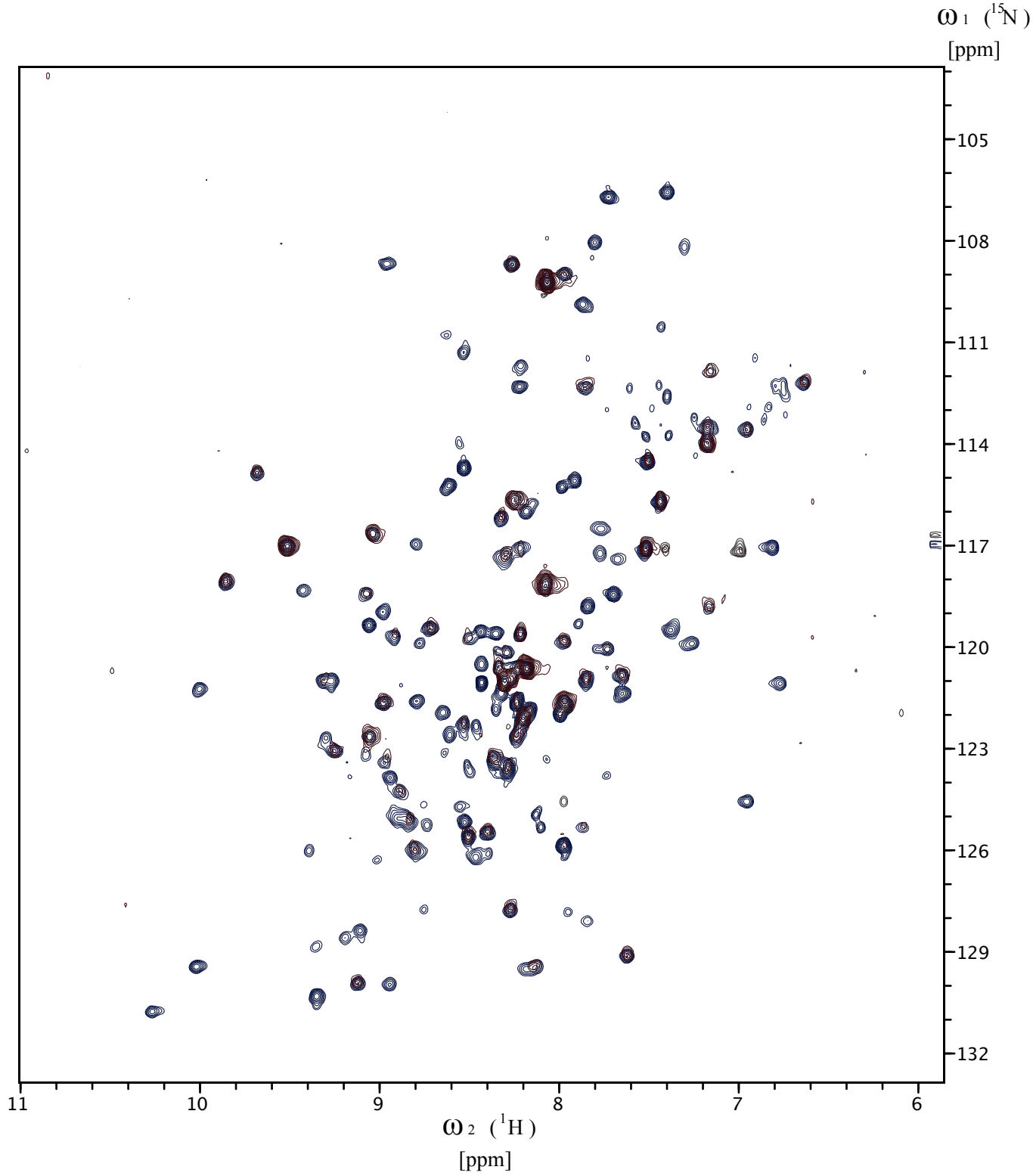
Appendix II : 2D HSQC of 0.39 mM L-PGDS in 20 mM HEPES pH6.5, 150 mM NaCl, 2 mM TCEP 10% D 2O, 0.5 mM DSS and 2.6 mM SA U44069



Appendix III : Backbone assignment of L-PGDS protein



Appendix IV : Overlay of ^{13}C Alanine labeled sample HSQC spectrum (blue) with HNCO spectrum (red).



Appendix V : Overlay of ^{13}C Leucine labeled HSQC spectrum (blue) with HNCO spectrum (red).